

Perspective-taking is spontaneous but not automatic



Cathleen O'Grady¹ , Thom Scott-Phillips^{2,3}, Suilin Lavelle¹ and Kenny Smith¹

Quarterly Journal of Experimental Psychology
2020, Vol. 73(10) 1605–1628
© Experimental Psychology Society 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1747021820942479
qjep.sagepub.com



Abstract

Data from a range of different experimental paradigms—in particular (but not only) the dot perspective task—have been interpreted as evidence that humans automatically track the perspective of other individuals. Results from other studies, however, have cast doubt on this interpretation, and some researchers have suggested that phenomena that seem like perspective-taking might instead be the products of simpler behavioural rules. The issue remains unsettled in significant part because different schools of thought, with different theoretical perspectives, implement the experimental tasks in subtly different ways, making direct comparisons difficult. Here, we explore the possibility that subtle differences in experimental method explain otherwise irreconcilable findings in the literature. Across five experiments we show that the classic result in the dot perspective task is not automatic (it is not purely stimulus-driven), but nor is it exclusively the product of simple behavioural rules that do not involve mentalising. Instead, participants do compute the perspectives of other individuals rapidly, unconsciously, and involuntarily, but only when attentional systems prompt them to do so (just as, for instance, the visual system puts external objects into focus only as and when required). This finding prompts us to clearly distinguish spontaneity from automaticity. Spontaneous perspective-taking may be a computationally efficient means of navigating the social world.

Keywords

Perspective-taking; dot perspective task; automaticity; spontaneity; directional orienting

Received: 12 March 2019; revised: 4 March 2020; accepted: 27 March 2020

Introduction

Everyday interactions with other people seem to require us to keep track of what those around us can see. Actions as simple as asking a friend to hand you an object, passing a football to a team member, or assessing whether an oncoming pedestrian has noticed your bicycle appear to require tracking what another individual can see—that is, visual perspective-taking. Taking the visual perspective of another individual is a form of mindreading, requiring a mental representation of another person's visual field (Apperly, 2011). However, it could be the case that behaviours like these are guided by a less complex cognitive process, such as directional orienting, in which an agent is simply aware of what appears in the direction that another individual is facing (Heyes, 2014). Currently, much debate on visual perspective-taking centres on the question of whether results in certain visual perspective-taking tasks are better explained by mentalising or by submentalising

processes such as directional orienting (Conway et al., 2017; Freundlieb et al., 2016, 2018; Gardner et al., 2018a, 2018b; Langton, 2018; Santiesteban et al., 2014; Zhao et al., 2015a).

One significant reason why these empirical issues are presently unresolved is methodological inconsistencies in the experimental literature. Despite the fact that much of the literature uses the same basic experimental task (see below),

¹School of Philosophy, Psychology & Language Sciences, University of Edinburgh, Edinburgh, UK

²Department of Cognitive Science, Central European University, Budapest, Hungary

³Department of Anthropology, Durham University, Durham, UK

Corresponding author:

Kenny Smith, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK.
Email: Kenny.Smith@ed.ac.uk

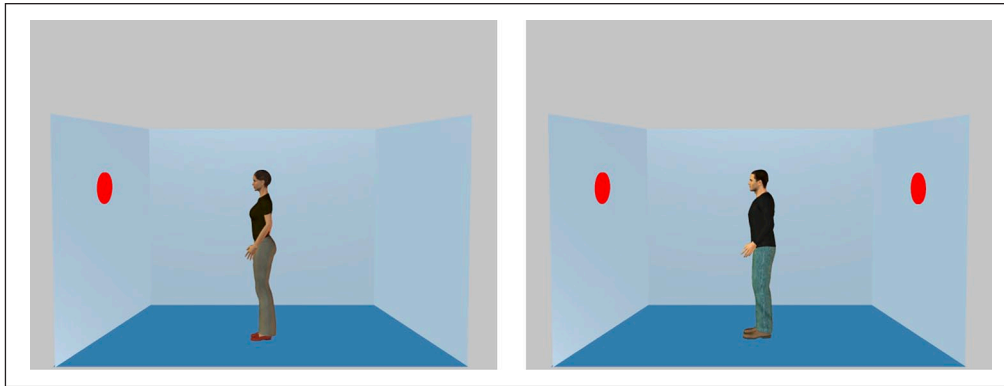


Figure 1. Stimuli from the original DPT (Samson et al., 2010). The task requires participants to view a scene that includes a human avatar and an array of dots. In every trial, participants are told whether to take the avatar's perspective (with the prompt HE or SHE) or their own perspective (with the prompt YOU). They are then shown a single digit in the middle of the screen, followed immediately by a scene such as those shown in this figure. They are asked to respond "Yes" or "No" depending on whether the digit matches the number of dots in the picture. On "Self" trials, participants must respond based on the number of dots they see in the picture. On "Other" trials, they must decide whether the digit matches the number of dots the avatar sees. The classic result in this paradigm is that participants react more slowly in *inconsistent* scenes (as pictured on the right), in which participants can see a different number of dots than the avatar, than in *consistent* scenes (left), in which they and the avatar can see the same number of dots (Bukowski et al., 2015; Samson et al., 2010; Santiesteban et al., 2014). This *consistency* effect occurs both when participants are reporting the number of dots the avatar can see (i.e., reaction times are slowed by the participant's own perspective; this is called *egocentric interference*) and when participants report their own perspective (this is called *altercentric interference*).

there are, nevertheless, recurrent variations in experimental design, making truly direct comparisons difficult. As we detail below, one crucial difference is the presence or absence of various prompts cueing participants to consider perspective-taking relevant to the task. This methodological choice is made for a variety of reasons, but especially key are differing assumptions about whether excluding prompts in certain tasks provides a more genuine assessment of spontaneous or automatic perspective-taking (Bukowski et al., 2015; Conway et al., 2017; Gardner et al., 2018a, 2018b; Santiesteban et al., 2014). There is further inconsistency in the use of the terms automatic and spontaneous themselves, which are used interchangeably in some papers, hindering clarity in the debate (Cole et al., 2016, 2017; Langton, 2018; Michael et al., 2018).

Here we address these issues. We first present a literature review that summarises the key issues identified above, discussing the utility of making a principled distinction between automatic and spontaneous processes. We then present three new preregistered studies that address the issues directly, using the same experimental task as much of the existing literature (the dot perspective task [DPT]; see below), and two replications using alternative stimuli. Collectively, our results show that one particular variant of the task does indeed demonstrate computation of another individual's perspective; that is, it involves perspective-taking rather than directional orienting. This effect arises rapidly and involuntarily (i.e., it is spontaneous), but it is not found uniformly across different task designs (i.e., it is not automatic). The effect depends instead on whether the perspective of the avatar

(or other stimulus) is made salient in one way or another. We further show that in another variant of the task, effects vary depending on the stimuli used, further corroborating the evidence that responses are not automatic, depending instead on participants' interpretation of the task requirements. Collectively, these results indicate that attentional processes moderate the deployment of perspective-taking. This finding explains apparent inconsistencies in the literature, and suggests that perspective-taking and directional orienting may both play a role in responses, depending on task context.

The DPT

The DPT requires participants to enumerate the number of dots that appear in a scene containing an avatar that sometimes has a different perspective from the participant's (see Figure 1 for a detailed description). The classic result is that participants are slower to respond based on their own perspective when the avatar's perspective differs from their own (Cole et al., 2016; Conway et al., 2017; Furlanetto et al., 2016; Nielsen et al., 2015; Qureshi et al., 2010; Samson et al., 2010; Santiesteban et al., 2014; Surtees & Apperly, 2012). This result is sufficiently well-established that in recent years the DPT has begun to be used to establish the presence or absence of perspective-taking abilities in a range of different contexts, including research on psychopathy and gender differences (Drayton et al., 2018; Yue et al., 2017).

However, the interpretation of results from the DPT is disputed. On one hand, data from the DPT are often cited

as evidence that participants “automatically” (Drayton et al., 2018; Furlanetto et al., 2016; Michael et al., 2018) or “spontaneously” (Cole et al., 2016, 2017; Gardner et al., 2018b; Samson et al., 2010; Surtees et al., 2016) compute the perspective of the avatar. This is because of the robust finding of altercentric interference: the conflicting perspective of the avatar slows down computation of what the participant herself sees. This occurs even on trials when the avatar’s perspective is strictly irrelevant to participants’ task of responding to the number of dots they (the participant) can see. Since computing the avatar’s perspective on these trials runs counter to the task instructions (both the instruction to take the perspective indicated on each trial, and the instruction to respond as rapidly as possible), and since the avatar’s perspective is not relevant to calculating the correct answer, the altercentric effect suggests that representation of the avatar’s perspective occurs involuntarily on these trials.

On the other hand, some variants of the DPT produce results that motivate an alternate explanation, namely that the altercentric interference effect is caused not by participants taking the perspective of the avatar and being slowed accordingly, but rather by the avatar serving as a directional cue directing participants’ attention to certain dots (Cole et al., 2016, 2017; Langton, 2018; Santiesteban et al., 2014). That is, altercentric interference may be explained not by participants forming a representation of the avatar’s line of sight, but rather by preferentially attending to the dots that the avatar “points” toward.

The following section discusses various versions of the DPT that have been used to investigate these issues, and the corresponding differences in task design that make the results from various studies difficult to reconcile.

Perspective-taking or directional orienting: differences in task design

An early modification to the DPT investigated whether altercentric interference would be found for stimuli that had a direction, but no agency of their own¹ (Santiesteban et al., 2014). This study found altercentric interference not only for avatars, but also for arrows, which was interpreted as evidence that avatars (and arrows) serve as a type of directional stimulus, prompting directional orienting rather than visual perspective-taking itself. There might, however, be different processes involved in each case: visual perspective-taking in the case of the avatar, and directional orienting in the case of the arrows (Cole et al., 2016). Indeed, gaze-cueing research suggests that, while eye gaze cues participants to a specific location, an arrow provides a more general cue (Marotta et al., 2012).

A second series of modified DPT variants instead manipulates what the avatar appears able to see, using either barriers that block the dots from the avatar’s field of view, or cartoon blindfolds or opaque goggles (Baker et al., 2016; Cole et al., 2016; Conway et al., 2017;

Furlanetto et al., 2016). We call these “occlusion” tasks. The idea here is that if altercentric interference is driven by directional orienting, then it should occur whenever the number of dots the avatar faces is lower than the overall number of dots in the scene, even if the avatar cannot “see” the dots (e.g., due to an occluding barrier or other method of blinding). If the effect is instead driven by perspective-taking, altercentric interference should not appear when the avatar is blinded, since the avatar cannot “see” any of the dots in either consistent or inconsistent scenes. These tasks have produced contradictory results, with some finding effects supportive of the perspective-taking account (Baker et al., 2016; Furlanetto et al., 2016) and others supporting the directional orienting account (Cole et al., 2016; Conway et al., 2017; Langton, 2018).

One possible explanation of these various contradictory results is that these experiments differ in whether participants are ever required to take the perspective of the avatar. Most of the experiments in the first, pioneering DPT study (Samson et al., 2010) required participants to answer based on their own perspective on some trials (“Self” trials), and based on the avatar’s perspective on others (“Other” trials). We call these *explicit* tasks, because the avatar’s perspective is explicitly relevant in these tasks. Explicit tasks can establish the presence of both egocentric and altercentric interference: on “Other” trials, explicit tasks may demonstrate *egocentric* interference, or slower judgements of the avatar’s perspective due to interference from one’s own perspective, since they are the only tasks that require participants to take the avatar’s perspective. On “Self” trials, explicit tasks may demonstrate *altercentric* interference, or slower judgements of one’s own perspective due to interference from the avatar’s perspective (see Figure 1).

This first DPT study (Samson et al., 2010) also included one task (Experiment 3) in which participants respond based only on their own perspective throughout the task. This experiment was motivated by concerns that mixing “Self” and “Other” trials may have cued participants to take the avatar’s perspective even on trials where it was not relevant (i.e., on “Self” trials). Participants were prompted with the cue “YOU” before every trial, and were instructed to ignore the central stimulus. We call tasks like this *implicit* tasks, because although they do not require participants to take the avatar’s perspective as part of the task, they do overtly mention the avatar and its perspective—whether to instruct participants to ignore the avatar’s perspective, as in Santiesteban et al. (2014), or to clarify for participants what the avatar can and cannot see, as in Cole et al. (2016). These instructions, along with the use of the word YOU as a cue on each trial, may still serve to prompt the participants to consider the avatar’s perspective as relevant to the task, hence the label *implicit*. Implicit tasks are capable of establishing only altercentric interference, not egocentric interference; but altercentric interference is the effect that drives the claim of automatic/

spontaneous perspective-taking, and so is the primary effect of interest in the DPT.

Note that the use of the terms “explicit” and “implicit” in this sense differ slightly from their use in the wider Theory of Mind literature, which distinguishes between explicit tasks that require a verbal response about another individual’s mental states, and implicit tasks that infer the presence of the representation of another individual’s mental states based on non-verbal responses (see, for example, San Juan & Astington, 2017). Here, we are using the terms to refer to the task instructions and demands; that is, to describe whether participants are *explicitly* or *implicitly* required to take the perspective of the avatar throughout the task.

Occlusion tasks have generally opted to use either an explicit design throughout a battery of tasks, or an implicit design throughout. Those using explicit tasks have tended to find evidence consistent with the perspective-taking account (Baker et al., 2016; Furlanetto et al., 2016), while those using implicit tasks have tended to find evidence consistent with directional orienting (Cole et al., 2016; Conway et al., 2017; Langton, 2018). One study has compared an explicit and implicit task, but this was done within subjects, in a substantially altered version of the DPT, making the findings difficult to interpret (Conway et al., 2017).

One further possibility is *uncued* tasks, which make no mention of perspective-taking in any of the information given to participants, have no requirement to take the avatar’s perspective, and no trial-by-trial “YOU” cue that could implicitly contrast the participant’s perspective with some other perspective. These tasks find no altercentric interference effect, unless there are further task modifications that draw additional attention to the avatar in some other way (for instance, having the avatar appear up to 600ms before the dots in the scene; Bukowski et al., 2015; Gardner et al., 2018b). (In the tasks that draw attention to the avatar in some way, results have been consistent with both the directional orienting and perspective-taking accounts, since these were not occlusion tasks. No existing uncued task attempts to discriminate between these.) In the Supplementary Information we describe a pilot study (uncued) that reports the same pattern of results.

In sum, apparently inconsistent results across variants of the DPT task may plausibly be due to differences in whether the perspective of the avatar—or other stimulus, such as an arrow—is made salient in one way or another, regardless of whether that perspective is strictly relevant for the task. This possibility prompts us to clearly distinguish between automatic and spontaneous cognitive processes, as described in the next section.

Implications for automaticity and spontaneity

Much of the experimental literature on the DPT is presented as informing the debate on “spontaneous perspective-taking” or “automatic perspective-taking.” These terms are

not often distinguished and sometimes used interchangeably. Few studies discuss exactly what spontaneity and/or automaticity entail. Where there is such discussion the most common approach is to say that for visual perspective-taking (or directional orienting) to be automatic or spontaneous, it should be purely stimulus-driven (Bukowski et al., 2015; Cole et al., 2016; Gardner et al., 2018b; Langton, 2018). That is, it should occur reflexively and mandatorily on seeing the avatar, without any cues to participants to take the avatar’s perspective, and without any need or motivation on the part of the participants to consider the avatar’s perspective relevant to the task (Cole et al., 2016; Gardner et al., 2018b; Langton, 2018). Whether these conditions are appropriate can be disputed. For instance, some researchers have suggested that automaticity is best conceived of not as a binary, but rather as a matter of degree, in which features such as goal-directedness, intentionality, control, and purely stimulus-driven response each play a partial role in establishing whether a process is automatic (Moors & De Houwer, 2006). Yet, the more narrow definition of automatic as purely stimulus-driven is fairly widespread in the DPT literature.

We suggest that *automatic* and *spontaneous* cognitive processes should be clearly distinguished (see also Carruthers, 2017; Westra, 2017). We consider automatic processes to be those that are reflexive and cannot be inhibited. In contrast, spontaneous processes are unconscious, involuntary, and rapid, but their operation is determined by intention, attention, or some other form of calibration. As an example of the difference, contrast seeing in colour, which is automatic, with seeing in focus, which is spontaneous: it occurs only as and when necessary, as determined by attention.

The varying empirical results reviewed above suggest two separate, but related, questions about visual perspective-taking:

1. Does the altercentric interference effect found in the DPT provide evidence of visual perspective-taking or directional orienting?
2. Does the process driving altercentric interference (whether visual perspective-taking or directional orienting) arise automatically, spontaneously, or neither?

The current literature suggests that the principal effect in the DPT is moderated by top-down appraisal of the task context (Bukowski et al., 2015; Gardner et al., 2018a, 2018b). In basic uncued tasks, with no awareness of the potential relevance of perspective-taking, there is no effect, while in uncued tasks when attention is drawn to the avatar in some way, there is an effect. In implicit tasks where there is minimal awareness of the presence of the avatars, there tends to be a directional orienting effect; visual perspective-taking effects occur only in explicit tasks, where there is a requirement to actively model the

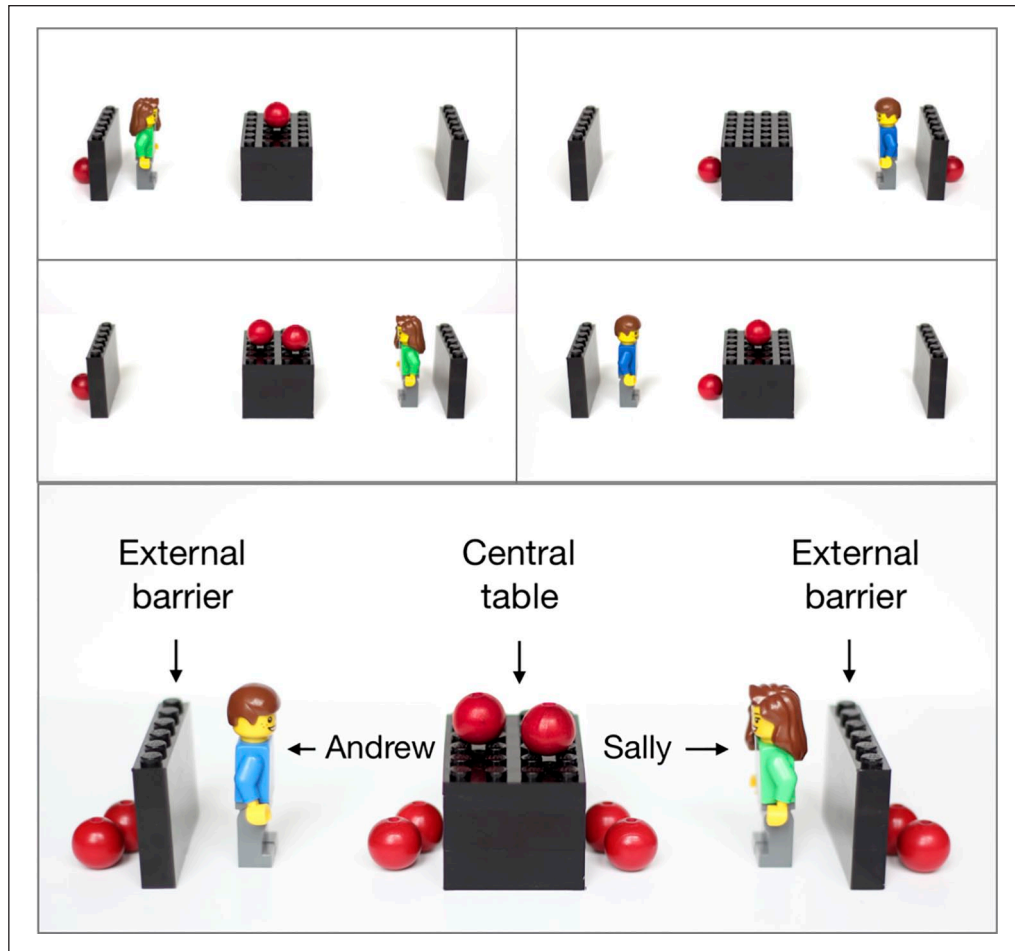


Figure 2. Adapted DPT stimuli using Lego figures. The upper four images show example scenes; note that each scene that participants saw featured a single avatar and a maximum of four balls. The lower image shows both potential placement positions for avatars (left or right of the central table) and all possible ball positions (five possible positions, a maximum of two balls in any one position).

perspective of the avatars. In explicit tasks, perspective-taking is voluntary at certain points during the task, but is nonetheless involuntary on those trials where the avatar's perspective is irrelevant to the immediate question. This pattern suggests that perspective-taking is not automatic, but may be spontaneous—that is, occurring rapidly and involuntarily on individual trials where the avatar's perspective is irrelevant, but only in an overall task where perspective-taking is relevant.

We present five experiments (three preregistered novel experiments, two replications using different stimuli) testing the hypothesis that the varying results reported in the literature are a consequence of task design. We first contrast explicit, implicit, and uncued versions of the DPT in a between-subjects design. Based on our reading of published results, we predicted that the explicit task would show an effect consistent with visual perspective-taking rather than directional orienting; that the implicit task would show directional orienting; and that the uncued task would show no effect. Findings matching these predictions

would suggest a continuum of attention to the avatar's perspective, depending on motivation created by task context, and that both visual perspective-taking and directional orienting arise spontaneously but not automatically. We then present a series of implicit tasks that attempt to establish the conditions under which an altercentric effect is found in the implicit condition.

Experiment 1: explicit, implicit, and uncued

Materials and methods

We constructed a new set of stimuli using photographs of Lego figures, dubbed “Sally” and “Andrew” for ease of reference (Figure 2).² We did this to increase task complexity for a planned series of experiments (not reported here) using multiple avatars simultaneously. Unlike the cartoon avatar used in most DPTs to date, these scenes had the benefit of unambiguous depth in the third

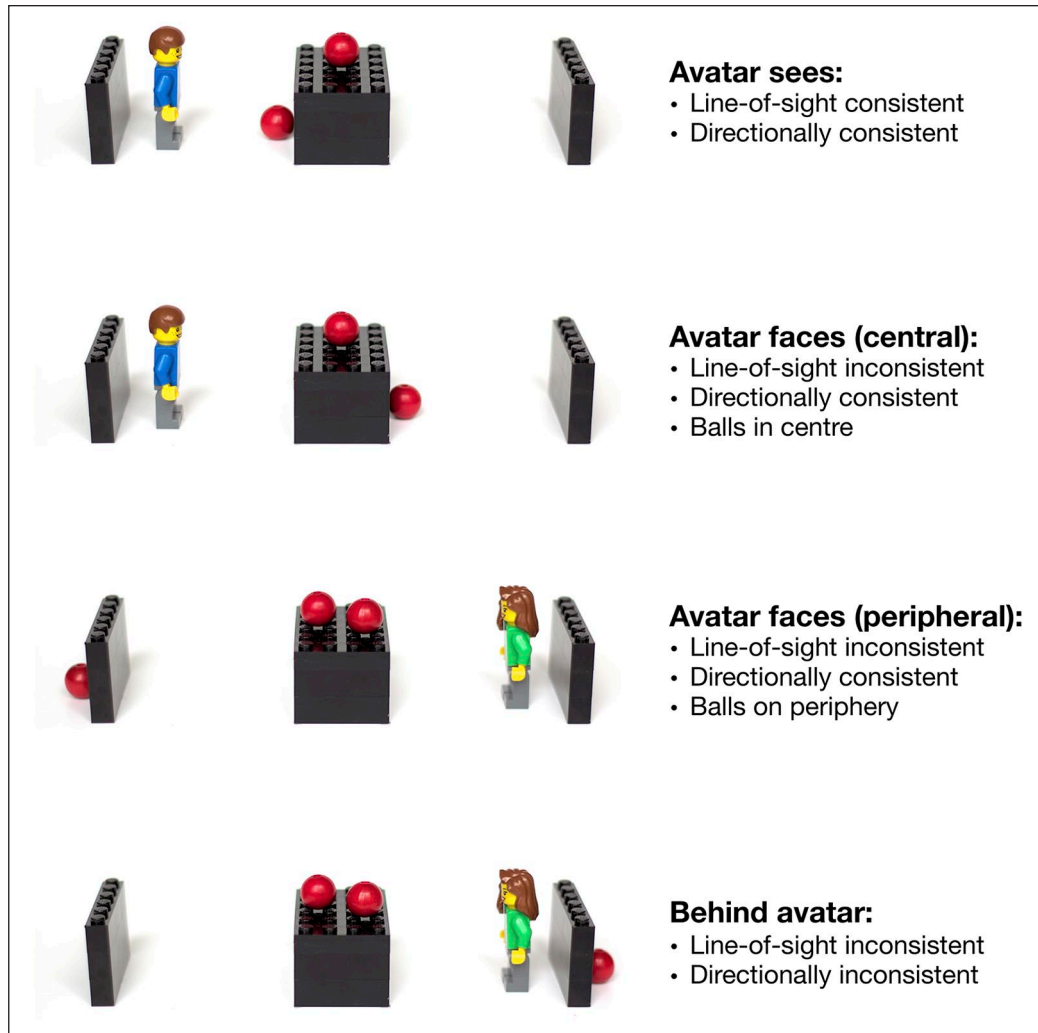


Figure 3. Example scenes from the four consistency conditions capturing the differences between avatar and participant perspectives, as well as spatial distribution.

dimension, and solid black barriers were used to prevent any ambiguity in whether or not Lego figures were able to see through them. A variety of hiding places allowed balls (our equivalent of dots/discs) to be hidden from view of the Lego figures, even when placed in front of them. Specifically, the balls could appear in any of five positions: on a central table, visible to either figure; on either side of the table, at the feet of the Lego figure, and within view only of the figure on that side of the table; or on either external boundary of the scene, behind an external barrier, within view of neither figure. Each scene featured a single Lego character, either Sally or Andrew. Each figure could appear on either side of the screen, along with zero to four balls and a maximum of two balls in any given location. The scenes were limited to four balls to allow for subitisation: that is, rapid and accurate enumeration of low numbers of items. Trick and Pylyshyn (1994) find that reaction times (RTs) remain low for subitisation of four items or fewer.

This layout allowed for two different definitions of perspective consistency (Figure 3). Line-of-sight consistency captures the inconsistent/consistent distinction used in the original DPT: line-of-sight consistent scenes are those in which there are no balls occluded from the avatar's perspective; the avatar and the participant can see the same number of balls. Line-of-sight inconsistent scenes are those in which the participant can see balls that are hidden from the avatar. A second definition of consistency describes whether the balls are in the direction that the avatar faces, regardless of whether or not they are occluded: directionally consistent scenes are those in which all balls are placed in the direction the avatar faces, while directionally inconsistent scenes are those in which balls appear behind the avatar.

Scenes may therefore be consistent by both definitions (*Avatar sees*), inconsistent by both definitions (*Behind avatar*), or line-of-sight inconsistent but directionally consistent (*Avatar faces*). Line-of-sight consistent, directionally

inconsistent scenes are not logically possible. Although previous research (Samson et al., 2010) controlled for the spatial layout of the room, confirming that the presence of the avatar and not merely the distance between the red dots was driving altercentric interference, it is possible that the greater complexity of our Lego scenes could introduce spatial artefacts. Specifically, scenes either have balls clustered entirely around the central table, or include balls on the periphery of the scene, outside the external walls. *Avatar sees* scenes are necessarily central; *Behind avatar* scenes are necessarily peripheral. Some *Avatar faces* scenes have balls only around the central table, while some include peripheral balls. *Avatar faces* scenes are, therefore, categorised further into *Avatar faces (central)*, allowing a comparison with *Avatar sees* scenes that controls for the spatial distribution of balls from the centre of the scene; and *Avatar faces (peripheral)*, allowing a spatial distribution-controlled comparison with *Behind avatar* scenes.

Based on our review of the DPT literature above, we made the following specific predictions for altercentric interference (i.e., from “Self” trials only):

1. Uncued, implicit, and explicit tasks will all result in slower RTs for scenes with dots positioned behind the avatar, compared with dots positioned in front of, and visible to, the avatar (i.e., *Behind avatar* vs. *Avatar sees* trials). There are three possible explanations for this effect: the spatial distribution of the scene, directional orienting, or visual perspective-taking. Further comparisons will discriminate between these possibilities.
2. The explicit task will show visual perspective-taking rather than directional orienting, illustrated by slower RTs on *Avatar faces (central)* than *Avatar sees* trials; that is, a delay when some balls are not visible to the avatar, even when they are in the direction that the avatar is facing. In the implicit and uncued conditions, we predict no difference between *Avatar faces (central)* and *Avatar sees* trials, suggesting no visual perspective-taking in these conditions.
3. The implicit and explicit tasks will show directional orienting, illustrated by slower RTs on *Behind avatar* than *Avatar faces (peripheral)* trials. That is, trials where all balls are in the direction the avatar is facing should be faster than those where balls are behind the avatar, suggesting directional orienting driving the *Behind avatar–Avatar sees* effect in the implicit task, and contributing to the effect in the explicit task. We expect that the uncued task will show no difference between *Behind avatar* and *Avatar faces (peripheral)* trials, suggesting that the *Behind avatar–Avatar sees* effect is driven purely by spatial distribution in this condition.

Preregistration. The experimental design and analysis was preregistered as part of the Open Science Framework’s Preregistration Challenge; the timestamped plan is available at <https://osf.io/5ey6d>.

Participants. Simulations based on a pilot experiment (see SI, Section 1) suggested that a sample size of 30 participants per condition would give substantially higher than 80% power at $\alpha = .05$ for the estimated effect sizes, for the within-subjects variables of interest. Ninety participants were recruited through the University of Edinburgh Student and Graduate Employment Service, and assigned randomly to the three between-subjects conditions: explicit, implicit, and uncued. They were compensated with £4 for their participation, which lasted approximately 30 min. Data were excluded from two participants whose tasks were interrupted by computer failure, and replaced by data from two new participants.³ Participants gave written consent, including consent for anonymised data to be shared publicly. Ethical approval was granted by the University of Edinburgh’s School of Philosophy, Psychology and Language Sciences Research Ethics Committee (PPLSREC), reference number 109-1718/1.

Procedure. On each trial, participants saw a fixation cross, followed by a one-word instruction, followed by a digit (0–4) presented for 750 ms, finally followed by a Lego scene accompanied by a prompt for a response. Figure 4 shows example trial sequences. In the explicit condition, participants were told that their task was to judge whether the digit they saw on each trial matched the number of balls that could be seen in the following picture. If they saw the word YOU before the trial (“Self” trials), they should answer based on how many balls they could see, and if they saw the word HE or SHE before the trial (“Other” trials), they should answer based on how many balls the Lego figure could see. In the implicit condition, participants were instructed to ignore what the Lego figure could see, and answer based only on what they could see. They were told that the word YOU would appear before each trial to remind them to answer based on their own perspective. In the uncued condition, participants were told that their task was to judge whether the digit matched the number of balls in the picture, with no mention of the Lego figure. The word READY? appeared before each trial, to make the trial length identical across conditions.

Participants completed a short training session explaining the task, followed by 32 practice trials (each followed by feedback informing the participant whether their answer had been correct or incorrect), and then the main task, divided into four blocks with self-paced breaks between blocks. On each trial, participants were presented with the cue word (YOU/HE/SHE/READY?, depending on condition) for 750 ms, followed by a fixation cross for 750 ms, and finally a digit between 0 and 4 for 750 ms,

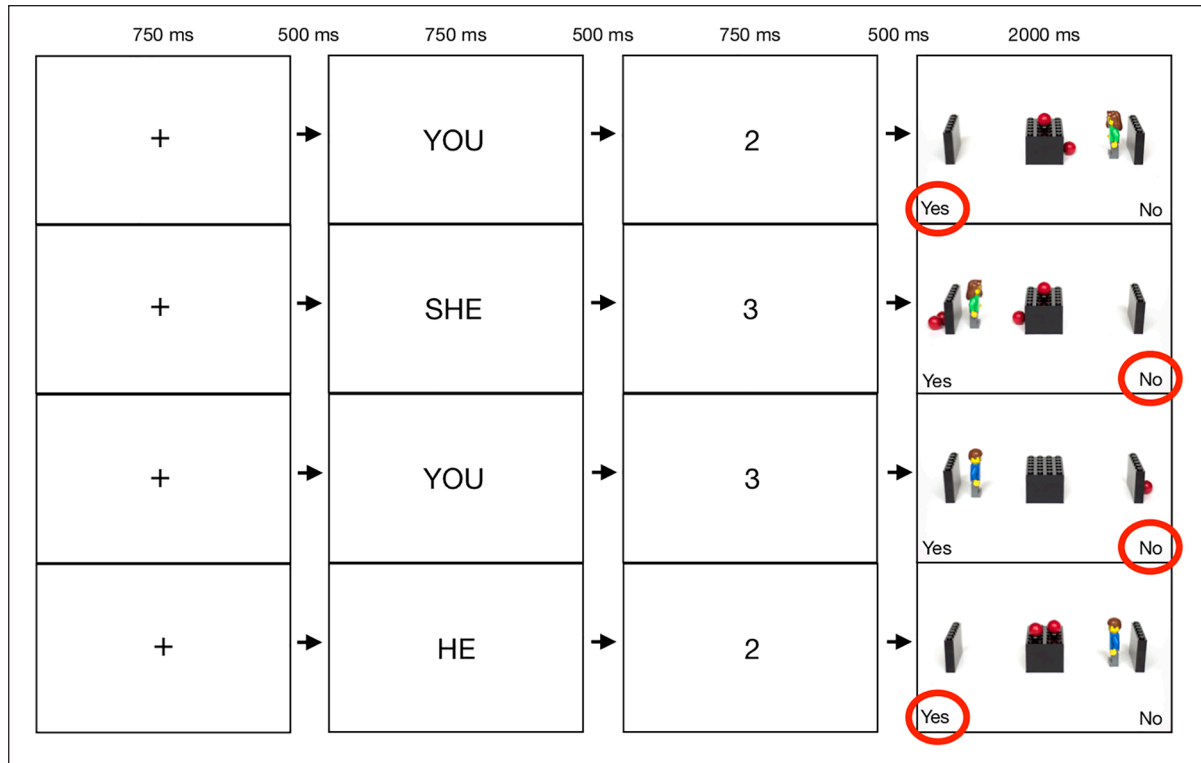


Figure 4. An illustration of the trial procedure in the explicit condition, with correct answers highlighted.

before the Lego scene appeared with the words “Yes” and “No” in the bottom corners of the screen. A two-button box was used to respond, with participants instructed to press the Yes-side button for yes and the No-side button for no. The “Yes” and “No” labels were presented on the screen to facilitate exact replication between tasks regardless of input equipment. The sides of these prompts were counterbalanced between participants, with half of the participants seeing “No” on the bottom left-hand corner of the screen throughout the task, and the other half seeing it on the bottom right-hand corner. This counterbalancing was done to avoid left-to-right reading bias possibly favouring the left-hand prompt, and the majority human left hemispheric dominance possibly favouring the right-hand prompt. Scenes timed out within 2,000 ms if no response was given, and the task moved on to the following trial.

The manipulated within-subjects variable of interest was the consistency between the avatar’s perspective and the participant’s perspective. For each participant, there were 64 trials in each of the four consistency conditions (*Avatar sees*, *Avatar faces [central]*, *Avatar faces [peripheral]*, and *Behind avatar*). In addition, a range of other constraints were followed, balancing which avatar appeared and on which side of the scene, the number of scenes with each possible number of balls, the number of yes versus no answers, and in the explicit condition, Self versus Other trials (see SI, Section 2).

The experiment was implemented using PsychoPy (Peirce, 2010).

Results

This design allowed the predictions detailed above to be tested using a series of mixed-effects models. All analyses reported below are in accordance with the preregistered analysis plan, unless otherwise noted.

We removed training trials, filler trials (those with zero balls), and timed-out trials (0.76%, $n = 176$); as per Whelan (2008), trials in which the response RT was lower than 100 ms were also removed, on the assumption that these trials could not be genuine responses to the stimuli (0.02%, $n = 5$). No trimming was conducted on higher RTs, given the imposed cut-off of 2,000 ms on all trials. Although Samson et al. (2010) and many subsequent DPT variants analyse only “Yes” trials on the basis that “No” trials may be easier to respond to, Santiesteban et al. (2014) found no difference between “Yes” and “No” responses. We therefore have not removed data from “No” trials and do not include this as a variable in our analyses. Because RT data deviates from the normal distribution, models used a log-transformed RT (logRT) as the dependent variable, to reduce skewing of the data and better conform to the assumptions of the model (Baayen & Milin, 2010).

A binomial logistic regression analysis of error rates in a pilot task (reported in the Supplementary Information)

Table 1. Results of Experiment 1, Model 1: *Avatar sees* versus *Behind avatar*.

Model	Slope	β	SE	χ^2	df	p
Main model	Consistency	.039	0.004	48.49	1	<.001***
	Consistency \times Condition (implicit vs. explicit)	.007	0.005	2.36	2	.31
	Consistency \times Condition (implicit vs. uncued)	-.005	0.004			
Planned comparisons	Consistency (explicit)	.044	0.008	22.62	1	<.001***
	Consistency (implicit)	.034	0.005	32.71	1	<.001***
	Consistency (uncued)	.032	0.006	24.70	1	<.001***

SE: standard error.

*** $p < .001$

failed to converge, presumably due to a lack of data, since error rates in the DPT are extremely low. We therefore removed trials with erroneous responses (4.2%, $n = 959$), but did not analyse them due to a lack of statistical power.

Because we were interested in altercentric interference, all “Other” trials were removed in the explicit condition (recall that the definition of implicit and uncued conditions is that there are no “Other” trials).⁴ This means that in all conditions, we are looking only at participants’ responses where they are evaluating whether the digit they were presented with matches the number of balls *they* can see. Any interference effects will therefore be altercentric interference, that is, due to inability to suppress the avatar’s perspective when that perspective is irrelevant on the trial at hand.

The explicit, implicit, and uncued tasks were of the same length to avoid differing fatigue effects between conditions, which halved the number of trials available for analysis in the explicit task: 3,607 explicit trials versus 7,235 implicit and 7,396 uncued.⁵ Given that all analyses were within-subjects in a particular condition and that the power analysis showed sufficient statistical power for this number of trials in the explicit condition, we see no reason that this could have accounted for any differences between conditions. We had no theoretically motivated predictions for Other trials, and these trials were therefore not analysed to limit researcher degrees of freedom in the analysis (see Simmons et al. [2011] for discussion of the problems associated with researcher degrees of freedom).

We used lme4 (Bates et al., 2015) and afex (Singmann et al., 2017) to perform a series of mixed-effects regression analyses on the logRTs. Mixed-effects models were used rather than the ANOVA used in previous experiments to avoid the necessity of averaging across observations for each participant, and to account for random effects—that is, the variance associated with different images as well as different participants. We used the standard $p < .05$ criterion for determining where effects were significant, with p values obtained using model comparison (likelihood ratio tests) using the mixed() function in the afex package (Singmann et al., 2017) in R (R Core Team, 2015).

Model 1: are Behind avatar scenes slower than Avatar sees scenes? We predicted that RTs in *Avatar sees* scenes (where

the avatar’s perspective matched the participant’s) and in *Behind avatar* scenes (where the avatar’s perspective mismatches the participant’s, under both line-of-sight or directional accounts) should differ (specifically, RTs in *Avatar sees* scenes should be faster) in all three tasks (explicit, implicit, and uncued), although possibly for different reasons, to be unpicked in subsequent analyses. To test this prediction, we ran an analysis on the data from *Avatar sees* and *Behind avatar* trials. Consistency and Condition (explicit vs. implicit vs. uncued) were sum-coded and entered as fixed effects, with interaction term, into the model. The sum coding for condition resulted in comparisons of explicit versus implicit, and implicit versus uncued. As differences in overall RT between the three conditions were not relevant to our predictions and had no theoretically motivated hypotheses about these differences, the results of these slopes are not reported. Random intercepts for images and participants were specified, as well as by-participant random slopes for the effect of consistency.⁶

The model (Table 1) showed an effect of Consistency, suggesting that *Behind avatar* trials were approximately 44.22 ms slower on average than *Avatar sees* trials. There was no interaction between Condition and Consistency, implying that all three conditions showed the same effect, with a 59.89 ms difference in the explicit condition, 38.93 ms in the implicit condition, and 35.66 ms in the uncued condition (Figure 5).

In all conditions, then, *Avatar sees* trials were associated with faster RTs than *Behind avatar* trials, matching our prediction. However, the cause of this effect (visual perspective-taking, directional orienting, or spatial distribution) is unclear.

Model 2: is there a mentalising effect in the explicit condition? We limited our data to *Avatar sees* and *Avatar faces (central)* trials—recall that in *Avatar faces (central)* trials, all balls in the scene are located centrally, but the participant and the avatar have distinct line-of-sight perspectives, that is, some balls are “hidden” from the avatar behind the central table. Otherwise, the model was identical to Model 1. The model (Table 2) showed no significant effect of Consistency, but a significant interaction between Condition and Consistency. Planned pairwise comparisons showed

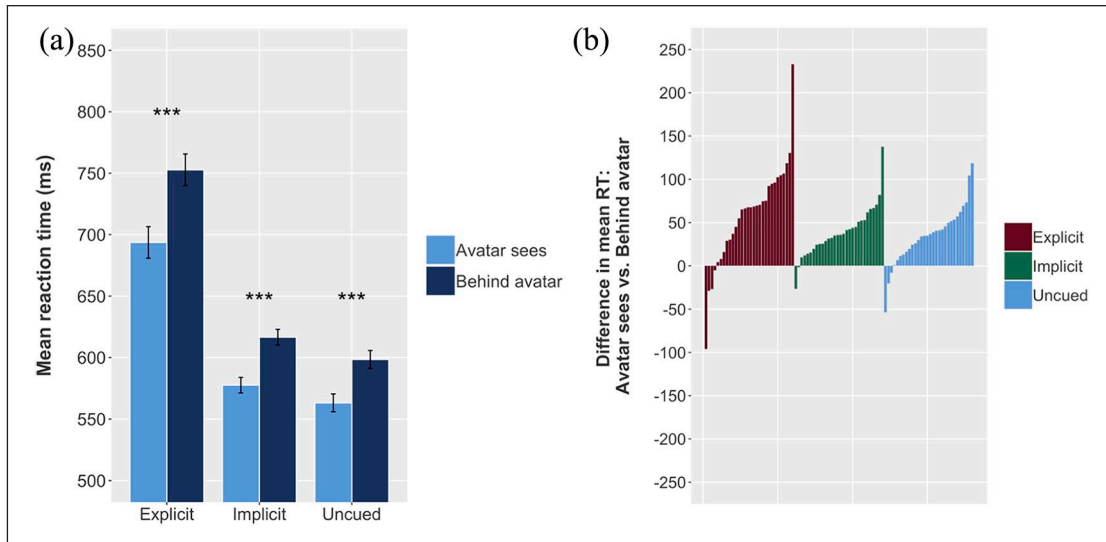


Figure 5. Effects of Experiment 1, Model 1: *Avatar sees* versus *Behind avatar*. (a) Mean RT for *Behind avatar* and *Avatar sees* conditions, for explicit, implicit, and uncued tasks; error bars indicate 95% CIs on the mean of the by-participant means, and significance annotations on the plots reflect the planned comparisons showing the effect of consistency for each condition. (b) Each individual participant's difference between mean *Behind avatar* RT and mean *Avatar sees* RT; lines extending above 0 on the y-axis indicate that the participant was slower in *Behind avatar* than in *Avatar sees* trials (i.e., exhibited an altercentric interference-like effect), while lines extending below 0 indicate that the participant was slower in *Avatar sees* than in *Behind avatar* trials. Mean reaction time is higher (i.e., participants respond more slowly) for *Behind avatar* trials in all three conditions (a); a substantial majority of participants in all three conditions show this effect (b).

Table 2. Results of Experiment 1, Model 2: *Avatar sees* versus *Avatar faces*.

Model	Slope	β	SE	χ^2	df	p
Main model	Consistency	.005	0.004	1.53	1	.22
	Consistency \times Condition (implicit vs. explicit)	.015	0.005	11.32	2	.003***
	Consistency \times Condition (implicit vs. uncued)	-.008	0.004			
Planned comparisons	Consistency (explicit)	.021	0.009	4.93	1	.03*
	Consistency (implicit)	-.005	0.004	1.10	1	.29
	Consistency (uncued)	-.003	0.005	0.48	1	.49

SE: standard error.

* $p < .05$

*** $p < .001$

that *Avatar faces* (*central*) trials were, on average, 27.79 ms slower than *Avatar sees* trials in the explicit condition, but showed no significant difference in the implicit or uncued conditions (Figure 6). This matches our prediction, and suggests visual perspective-taking in the explicit condition, and either a directional orienting or a spatial-distribution effect underlying the results for the implicit and uncued conditions in Model 1.

Model 3: is there a directional orienting effect in the implicit condition? We limited our data to *Avatar faces* (*peripheral*) and *Behind avatar* trials—recall that in *Avatar faces* (*peripheral*) trials the participant and the avatar have distinct line-of-sight perspectives, that, some balls are “hidden” from the avatar in a peripheral position, in the direction that the avatar is facing but behind one of the outer barriers;

in *Behind avatar* trials some balls are “hidden” behind the avatar, again in a peripheral position. The model (Table 3, model structure and coding as per previous analyses) showed no effect of consistency and no interaction between condition and consistency.⁷ Planned pairwise comparisons showed a significant effect for explicit but not implicit or uncued conditions. Note, however, that given the omnibus model showed no interaction, the significant effect (24.24ms) in the model analysing the explicit condition only should be treated with caution (Figure 7).

There is therefore (somewhat tentative) evidence matching our predictions for the explicit condition (where we expected a directional orienting component to visual perspective-taking, here indicated by participants responding faster when balls were in the direction the avatar was facing, even though they were occluded from the avatar's line of

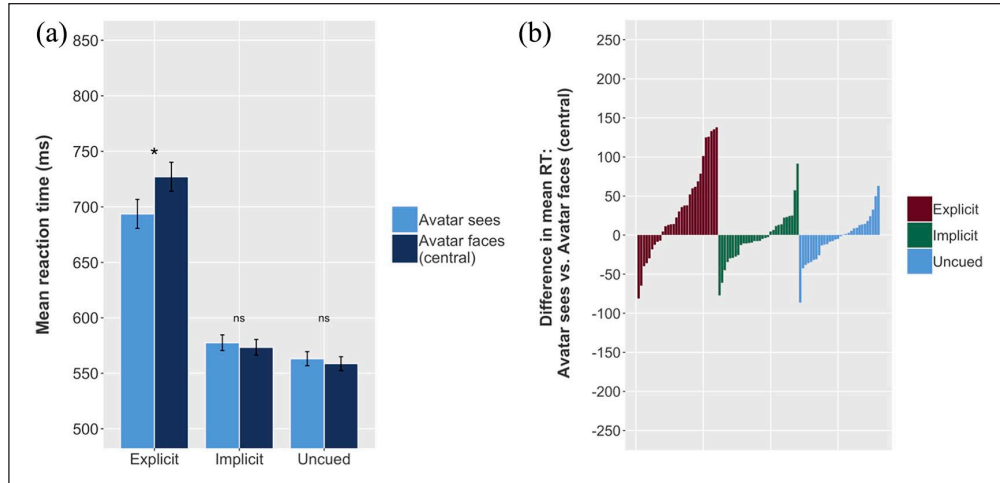


Figure 6. Effects of Experiment 1, Model 2: *Avatar sees* versus *Avatar faces*. (a) Mean RT for *Avatar faces* (central) and *Avatar sees* conditions, for explicit, implicit, and uncued conditions; error bars indicate 95% CIs on the mean of the by-participant means. (b) Each individual participant’s difference between mean *Avatar sees* RT and mean *Avatar faces* (central) RT. Mean reaction time is higher (i.e., participants respond more slowly) for *Avatar faces* (central) trials in the explicit condition, but not in the implicit or uncued conditions (a); a substantial majority of participants in the explicit condition, but not in the implicit or uncued conditions, show this effect (b).

Table 3. Results of Experiment 1, Model 3: *Avatar faces* (peripheral) versus *Behind avatar*.

Model	Slope	β	SE	χ^2	df	p
Main model	Consistency	.008	0.004	3.92	1	.05
	Consistency \times Condition (implicit vs. explicit)	.009	0.005			
	Consistency \times Condition (implicit vs. uncued)	-.005	0.004	3.90	2	.14
Planned comparisons	Consistency (explicit)	.018	0.007	5.45	1	.02*
	Consistency (implicit)	.004	0.006	0.45	1	.50
	Consistency (uncued)	.004	0.005	0.58	1	.44

SE: standard error.

*p < .05

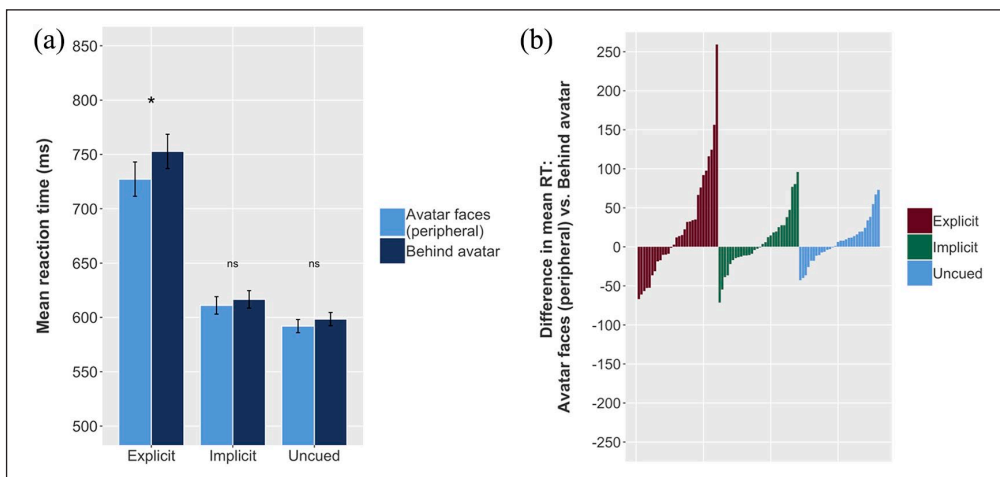


Figure 7. Effects of Experiment 1, Model 3: *Avatar faces* (peripheral) versus *Behind avatar*. (a) Mean RT for *Behind avatar* and *Avatar faces* (peripheral) conditions, for explicit, implicit, and uncued conditions; error bars indicate 95% CIs on the mean of the by-participant means. (b) Each individual participant’s difference between mean *Behind avatar* RT and mean *Avatar faces* (peripheral) RT. Mean reaction time is higher (i.e., participants respond more slowly) for *Avatar faces* (peripheral) trials in the explicit condition, but not in the implicit or uncued conditions (a); a small majority of participants in the explicit condition, but not in the implicit or uncued conditions, show this effect (b).

sight). These results also match our prediction for the uncued condition, where we predicted no effect of the avatar, and a *Behind avatar–Avatar sees* altercentric interference effect (see Model 1) driven entirely by central versus peripheral distributions of the balls. However, we do not find evidence matching our prediction for the implicit condition, where we predicted altercentric interference in this model, driven by directional orienting. Rather, our results suggest that our implicit and uncued conditions behave similarly, with the only effects we see being driven by the spatial distribution of the balls, with slower responses to scenes featuring balls in the periphery of the scene.

Discussion

These results support our hypothesis that the requirement to take the avatar's perspective on some trials results in visual perspective-taking; and that differences in task design or framing explain apparently conflicting results in the literature. We can manipulate the presence/absence of an altercentric interference effect by switching between an explicit task and implicit or uncued tasks.

In the explicit task, *Avatar sees* trials were 59.89 ms faster than *Behind avatar* trials, suggesting a spatial, perspective-taking, or directional orienting effect, or some combination of the three; *Avatar sees* trials were 27.79 ms faster than *Avatar faces (central)* trials, suggesting perspective-taking; and *Avatar faces (peripheral)* trials were 24.24 ms faster than *Behind avatar* trials, suggesting directional orienting. The considerably larger effect in *Behind avatar* versus *Avatar sees* suggests that the processes may be summative; that is, with both the distribution of balls from the centre of the scene and the avatar's perspective causing individual delays that result in a larger overall delay. The evidence for directional orienting (although this evidence is tentative, given the lack of omnibus effect in this model) suggests that directional orienting may play a role in perspective-taking, or otherwise operate in tandem with it, perhaps as a first visual sweep of a scene. The precise interaction of these varying effects would be a useful subject for future research.

The results are coherent with previous research using explicit and uncued tasks, but conflict with several studies that find altercentric interference in implicit tasks (Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014), likely driven by directional orienting (Cole et al., 2016; Conway et al., 2017).

A potentially important difference between our Lego stimuli and standard DPT stimuli is the positioning of the avatar. Previous implicit tasks (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014) have placed the avatar in the centre of the screen, preceded by a fixation cross and trial-by-trial instructions in the centre of the screen. Our stimuli position the avatar off-centre, preceded by the fixation cross and trial-by-trial

instructions in the centre of the screen. Given the literature suggesting that additional attention drawn to the avatar induces an altercentric effect even on uncued tasks (Bukowski et al., 2015; Gardner et al., 2018b), it is possible that previous implicit tasks have drawn additional attention to the avatar through the placement of the fixation cross and instructions over the spot where the avatar will appear (see, for example, Bukowski et al., 2015).

We therefore conducted a second preregistered implicit task, identical to the implicit condition in Experiment 1 but with the fixation cross and trial-by-trial instructions (i.e., the text "YOU" and the digit to be confirmed) placed directly over the point on the screen where the avatar will appear. We predicted that we would find the expected *Avatar faces (peripheral)* versus *Behind avatar* altercentric interference in this condition. This would suggest that attention must be drawn directly to the avatar on a trial-by-trial basis to induce directional orienting, implying that neither visual perspective-taking nor directional orienting is automatic; rather, they appear in response to ongoing cues regarding the avatar's relevance to the task.

Experiment 2: salience of avatars

Materials and methods

The same stimuli were used as for Experiment 1.

Preregistration. The experimental design and analysis was preregistered as part of the Open Science Framework's Preregistration Challenge; the timestamped plan is available at <https://osf.io/dk86n>.

Participants. Simulations based on Experiment 1 suggested that a sample size of 30 participants per condition would give substantially higher than 80% power at $\alpha = .05$ for the estimated effect sizes, for the within-subjects variables of interest. Thirty further participants were recruited through the University of Edinburgh Student and Graduate Employment Service. They were compensated with £4 for their participation, which lasted approximately 30 min. Data were excluded from one participant whose task was interrupted by computer failure, and replaced by data from a new participant. Participants gave written consent, including consent for anonymised data to be shared publicly. Ethical approval was granted by the University of Edinburgh's PPLSREC, reference number 188-1718/1.

Procedure. This task used the same procedure and task design as the implicit condition in Experiment 1. Fixation crosses and pre-trial instructions (the appearance of the word YOU and the digit between 0 and 4) were changed to appear centred over the position in which the face of the Lego character would appear in the following scene, rather than appearing centrally on the screen.

Table 4. Results of Experiment 2.

Slope	β	SE	χ^2	df	p
Behind avatar versus Avatar sees	.028	0.005	23.72	1	<.001***
Avatar faces (central) versus Avatar sees	-.001	0.005	0.07	1	.80
Behind avatar versus Avatar faces (peripheral)	.001	0.006	0.04	1	.84

SE: standard error.

***p < .001

Results

We applied the data exclusions and transformations described in Experiment 1, removing timed-out trials (0.49%, $n = 38$) and erroneous responses (3.36%, $n = 257$). There were no responses below 100 ms.

Following our preregistered analysis plan, data limited to the three relevant comparisons (*Behind avatar* vs. *Avatar sees*, *Avatar faces [central]* vs. *Avatar sees*, and *Avatar faces [peripheral]* vs. *Behind avatar*) were modelled using three models identical to the pairwise comparisons for the implicit condition in Experiment 1.

As in Experiment 1, and as predicted, the models showed a significant difference between *Avatar sees* and *Behind avatar* (35.96 ms), and no significant difference between *Avatar faces (central)* and *Avatar sees* trials (-1.55 ms, see Table 4). However, contrary to our predictions, there was also no significant difference between *Avatar faces (peripheral)* and *Behind avatar* trials, at 1.43 ms (Figure 8). This suggests that there was no directional orienting effect in this task, and that the difference between *Avatar sees* and *Behind avatar* trials was driven by the spatial distribution of the balls.

Discussion

These results do not support the hypothesis that directional orienting played any role in this implicit task. These results continue to conflict with findings of consistency effects in implicit tasks (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014).

One possible explanation for this could be the complexity of the scene. The original DPT used a simple scene consisting only of the avatar in a room, with an array of dots. Occlusion tasks have used a single avatar that appeared in a consistent position, with up to three balls and one or two barriers (Baker et al., 2016; Cole et al., 2016) or another method of blinding that added a single element to the existing scene, such as goggles or a telescope (Conway et al., 2017; Furlanetto et al., 2016). It may be that the Lego stimuli, with three barriers, two possible avatars appearing in two different places, and up to four balls, increased the scene complexity to the extent that participants' strategies to complete the task changed substantially. That is, it may be the case that participants were best able to respond quickly and accurately to each trial by

ignoring the perspective of the on-screen character—a strategy that would not be possible in an explicit task (explaining the results in Experiment 1) but would be possible in implicit and uncued tasks.

To explore the possibility of scene complexity driving the null results in these implicit tasks, we conducted a further preregistered implicit task, simplifying the Lego stimuli to scenes equivalent to those in the original DPT. These scenes consisted of a central figure, with balls level with the character's gaze, positioned either in front of or behind the character. Based on our reading of the extant literature, we predicted that scene complexity would explain the lack of altercentric effect on implicit tasks in Experiments 1 and 2, that is, that there would be an altercentric effect with simplified stimuli.

Because these simplified stimuli do not incorporate any barriers that distinguish between *Avatar sees* and *Avatar faces*, they are not able to provide evidence for whether any altercentric effect found in this task is better explained by perspective-taking or by directional orienting. However, the results of this task should help to explain the unexpected null results for directional orienting in the implicit task in Experiment 1, and in Experiment 2.

Experiment 3: reducing the visual complexity of the scene

Materials and methods

The images used in Experiment 1 and 2 were digitally edited to match the layout of the original DPT stimuli (Figure 9). Each Lego character appeared centrally on the screen, facing either left or right, with up to two balls in each scene. The balls, which floated at the height of the gaze of the Lego character, could appear in front of the character, behind it, or both in front and behind. As in Experiment 2, participants were instructed to ignore the perspective of the Lego character, and the word YOU appeared before each trial. The fixation cross and pretrial instructions appeared over the position where the face of the Lego character would appear.

Preregistration. The experimental design and analysis was preregistered as part of the Open Science Framework's Preregistration Challenge; the timestamped plan is available at <https://osf.io/hr98w>.

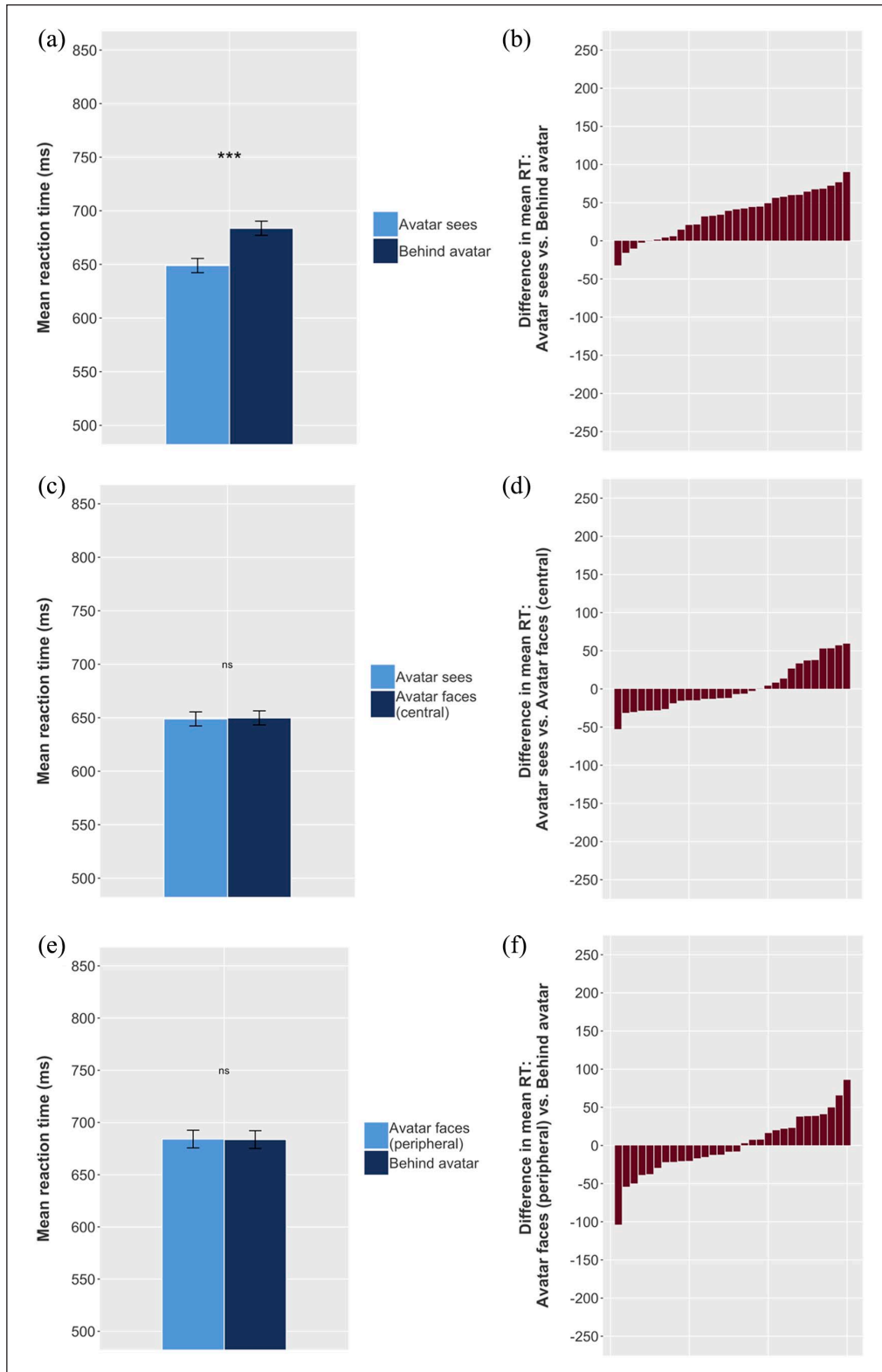


Figure 8. Results of Experiment 2. (a) Mean RT for *Behind avatar* and *Avatar sees* conditions; error bars indicate 95% CIs on the mean of the by-participant means. (b) Each individual participant’s difference between mean *Behind avatar* RT and mean *Avatar sees* RT. Mean reaction time is higher (i.e., participants respond more slowly) for *Behind avatar* trials (a); a majority of participants show this effect (b). However, there is no difference in RT between (c, d) *Avatar sees* and *Avatar faces (central)* or (e, f) *Behind avatar* and *Avatar faces (peripheral)*.

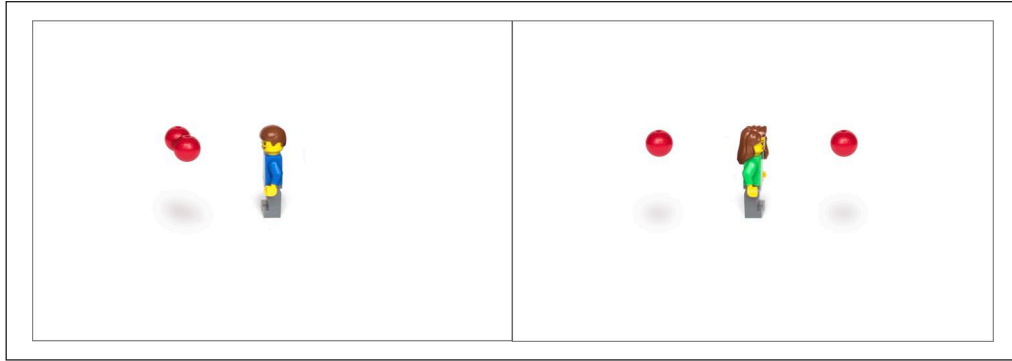


Figure 9. Lego stimuli adapted to match original DPT scene layout. Each scene consists of a single avatar and up to two balls, which can appear in front of or behind the avatar.

Participants. Sample size calculation was based on the same simulation method as Experiment 2. Thirty participants were recruited through the University of Edinburgh Student and Graduate Employment Service. They were compensated with £ 4 for their participation, which lasted approximately 30 min. Participants gave written consent, including consent for anonymised data to be shared publicly. Ethical approval was granted by the University of Edinburgh's PLSREC, reference number 188-1718/1.

Procedure. The procedure for Experiment 2 was used, with some differences. Participants completed 16 practice trials, followed by 192 test trials: 96 in which the avatar could see the same number of balls as the participant (*Avatar sees*) and 96 in which at least one ball was concealed behind the avatar (*Behind avatar*). Up to two balls appeared in any given scene. Avatar, yes/no response, and number of balls were balanced across trials (see SI, Section 3).

Results

We applied the data exclusions and transformations described in Experiment 1, removing timed-out trials (0.30%, $n=17$), erroneous responses (2.25%, $n=129$), and the single trial with a response below 100 ms (0.02%).

Following our preregistered analysis plan, a mixed-effects regression was used to compare the logRTs for *Avatar sees* trials with *Behind avatar* trials. Consistency was sum-coded and entered as a fixed effect, and random intercepts for images and participants were specified, as well as by-participant random slopes for the effect of Consistency. Contrary to our prediction, the model showed no difference between *Avatar faces* and *Behind avatar* trials, at an estimated -2.60 ms (Table 5, Figure 10). This suggests (alongside the results of Experiments 1 and 2) an absence of any directional orienting in our implicit DPT. This is contrary to the findings of several existing studies (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014).

Table 5. Results of Experiment 3.

Slope	β	SE	χ^2	df	p
<i>Behind avatar</i> versus <i>Avatar sees</i>	-.002	0.003	0.43	1	.51

SE: standard error.

Discussion

This task found no evidence of difference in RT between *Behind avatar* and *Avatar sees* in an implicit task, contrasting with the results of Experiments 1 and 2. This contrast may be explained by differences in the spatial distribution of the balls: in Experiments 1 and 2, *Avatar sees* trials all had balls clustered in the centre of the screen, while *Behind avatar* trials had balls on the periphery of the scene. In Experiment 3, these two conditions had balls evenly placed from the centre of the screen. The lack of effect in Experiment 3, with balls evenly distributed from the centre of the scene in these two conditions, therefore contributes to the evidence that this effect was driven by spatial distribution in Experiments 1 and 2.

The difference between the null result in Experiment 3 and the altercentric effect found in several implicit tasks (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014) raises the possibility that there is an important difference between the Lego stimuli and the avatars used in previous tasks. While we find this surprising, it may be due to unanticipated differences in the willingness of participants to accept Lego avatars versus cartoon avatars as having a perspective. Given that many participants are likely to have interacted with Lego characters as objects, but all would encounter the avatars for the first time in the context of the experiment, one possibility is that participants are more inclined to consider the Lego characters as objects but the cartoon-like avatars in the standard DPT stimuli as agents. The greater realism of the standard avatars may also contribute to a heightened perception of agency. We therefore ran a second simplified task using the original DPT stimuli, otherwise

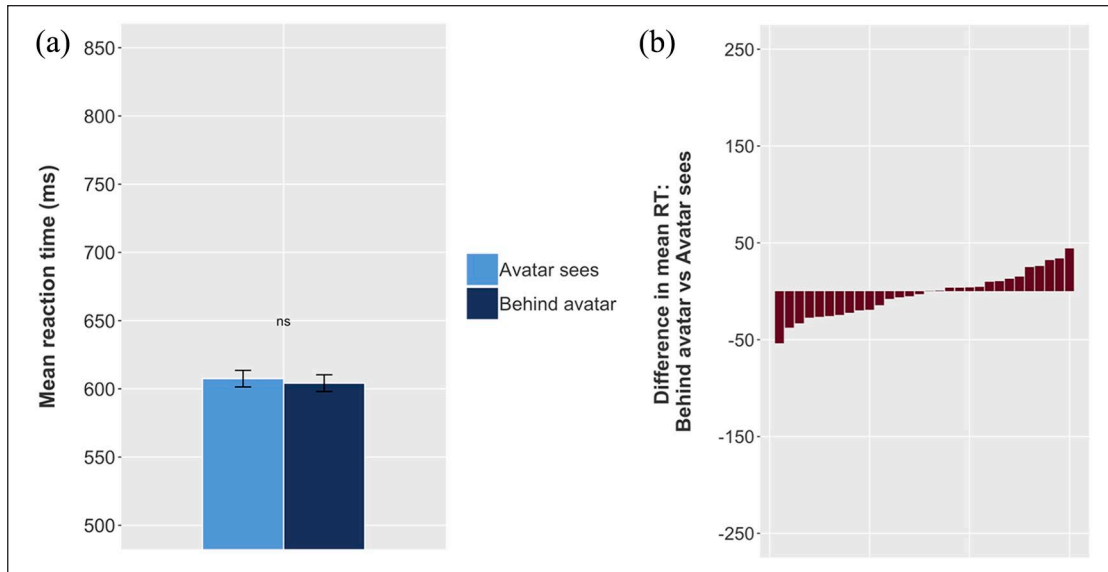


Figure 10. Results of Experiment 3. (a) Mean RT for *Behind avatar* and *Avatar sees* conditions; error bars indicate 95% CIs on the mean of the by-participant means. (b) Each individual participant's difference between mean *Behind avatar* RT and mean *Avatar sees* RT. Mean reaction time is not significantly different between the two conditions.

identical to Experiment 3. Because this was simply a replication of Experiment 3 using different stimuli, it was not preregistered separately, as all other details of the preregistration were the same.

Experiment 4: original stimuli

Materials and methods

The materials and methods for Experiment 3 were reused, with original DPT stimuli instead of Lego stimuli. The images were sized so that the on-screen characters were of the same height, and the characters' heads in the same position on the screen, as the Lego characters in Experiment 3.

Participants. Thirty participants were recruited through the University of Edinburgh Student and Graduate Employment Service. They were compensated with £ 4 for their participation, which lasted approximately 30 min. Participants gave written consent, including consent for anonymised data to be shared publicly. Ethical approval was granted by the University of Edinburgh's PPLSREC, reference number 188-1718/1.

Procedure. This task used the same procedure and task design as Experiment 3.

Results

We applied the data exclusions and transformations described in Experiment 1, removing timed-out trials (0.24%, $n = 14$) and erroneous responses (2.98%, $n = 171$). There were no responses below 100 ms.

Table 6. Results of Experiment 4.

Slope	β	SE	χ^2	df	p
<i>Behind avatar</i> versus <i>Avatar sees</i>	.010	0.004	7.05	1	.008**

SE: standard error.

** $p < .01$

The data were analysed using a model identical to that used in Experiment 3. The results showed a significant difference between *Behind avatar* and *Avatar sees* trials, at 11.30 ms (Table 6). Note that this is a substantially smaller effect than other implicit tasks using these stimuli: 21 ms (Samson et al. [2010], Experiment 3); 35.4 ms (Santesteban et al. [2014], Experiment 2); approximately 40 ms (Cole et al., 2016); 35 ms (Conway et al. [2017], Experiment 1). We conducted a further exploratory model comparing RTs across the two experiments, with Consistency and Stimulus entered as fixed effects (with interaction term) and the same random effects structure as the basic model. This model revealed a significant Consistency \times Stimulus interaction, providing further evidence of a consistency effect with the original stimuli, but not with the Lego stimuli (Table 7, Figure 11).

Discussion

These results suggest that, remarkably, the stimuli themselves play a role in producing an altercentric effect. The lack of an effect in the implicit tasks in Experiments 1–3 appears to be due to some difference between the Lego stimuli and the original stimuli, suggesting that the Lego stimuli do not result in either directional orienting or perspective-taking without additional direction to take the character's

perspective. It is possible that there are features of the scenes other than the avatars themselves driving this difference (for instance, the brightness of the colours; the overlap of balls in the Lego scenes compared with the spacing of the spacing of discs in the original stimuli; or the lack of a blue background room in the Lego scenes). A reviewer suggests that an alternative explanation is a difference between the directional features of Lego and cartoon avatars. That is, the original avatars may provide more cues for the front and back of the body compared with the Lego avatars: they have torsos with a clear front and back shape, and faces with humanoid profiles, compared with the flat faces and body shape of the Lego pieces. As non-humanoid stimuli have resulted in an altercentric effect (Santiesteban et al., 2014) in an implicit task, the directional cueing properties of the stimuli may play an important role.

Yet another explanation could be differences in attribution of agency to the Lego avatars compared with the cartoon avatars. This could be due to participants' experience of Lego characters as non-agentive objects in the real world, although Lego figures have been shown to be processed as animate in at least some circumstances (LaPointe

et al., 2016), or it may be due to intrinsic properties of the images—that is, the greater realism of the cartoon avatars, with near-human proportions, body shape, and facial projections.

Altercentric interference appears to be a robust effect in a wide range of simple DPT variants, and has even been found in more complex scene layouts with non-standard avatars (Baker et al., 2016; Mattan et al., 2015). The unexpected lack of altercentric interference in Experiments 1–3 can nonetheless be reconciled with the wider literature. The DPT variants that have used non-standard avatars have been explicit, and our Experiment 1 using Lego figures suggests that an explicit task may be sufficient to drive perspective-taking. Implicit tasks make up a limited sub-section of the DPT literature, and all use the standard avatar, or the standard avatar with minor modifications such as a blindfold or barrier (Cole et al., 2016; Conway et al., 2017; Samson et al., 2010; Santiesteban et al., 2014). There are two notable exceptions. First, Langton (2018) uses photographs of people in an implicit occlusion task, finding results consistent with directional orienting. This is coherent with both explanations discussed above; that is, that a photograph of a human would provide greater directional cues or clearer evidence of agentiveness than Lego figures in the same way that humanoid avatars would. Langton (2018) also uses live human experimenters in an uncued task that has a substantial delay between the orientation of the experimenter and the appearance of dots; as this is analogous to other uncued tasks that manipulate stimulus onset asynchrony (Bukowski et al., 2015; Gardner et al., 2018b), the finding of directional orienting in this task is not surprising. Second, Santiesteban et al. (2014)

Table 7. Lego versus original stimuli.

Slope	β	SE	χ^2	df	p
Original versus Lego	-.011	0.027	0.18	1	.67
Behind avatar versus Avatar sees	.004	0.003	2.34	1	.13
Interaction	.006	0.003	5.83	1	.02*

SE: standard error.

* $p < .05$

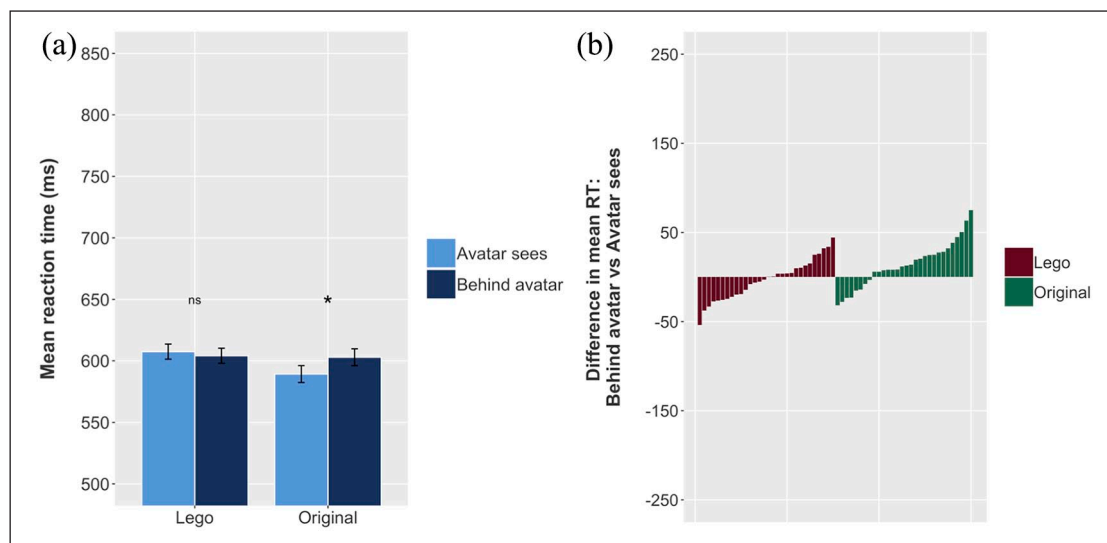


Figure 11. Comparison of effects in Experiments 3 and 4. (a) Mean RT for *Behind avatar* and *Avatar sees* conditions for both Lego and original stimuli; error bars indicate 95% CIs on the mean of the by-participant means. (b) Each individual participant's difference between mean *Behind avatar* RT and mean *Avatar sees* RT for both Lego and original stimuli. Mean reaction time is higher (i.e., participants respond more slowly) for *Behind avatar* trials for the original stimuli only (a); a majority of participants in this condition show this effect (b).

find an altercentric effect on an implicit task using arrows as stimuli; as discussed above, this is consistent with the explanation that sufficient directional cueing in a stimulus may be sufficient to trigger directional orienting.

These results contribute to the evidence suggesting that, while the altercentric effect may be widely replicated, it is nonetheless surprisingly sensitive to small differences in task design that prompt attention to the avatar and the relevance of its perspective. Prompts hinting at the relevance of certain kinds of agentive stimuli, such as discussion of the avatar's perspective and the inclusion of the YOU cue on every trial, may produce the altercentric effect in implicit tasks (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014), while other measures drawing attention to the avatar, such as the appearance of the avatar before the dots, may achieve the same effect (Bukowski et al., 2015; Gardner et al., 2018b). The results from Experiment 4 reported here suggest that the perception of agency of the avatar may be an alternative method of drawing attention to the avatar. They also provide further evidence against the automaticity of either a perspective-taking or directional orienting effect in the DPT, but combined with the results of an explicit task found in Experiment 1, suggest that ongoing attention drawn to the avatar leads to a rapid, involuntary (spontaneous) perspective-taking effect.

The simplified scene design used in Experiment 4 makes it impossible to determine whether the altercentric effect we observe here represents perspective-taking or directional orienting. This distinction requires an occlusion task, and these results suggest that the implicit occlusion tasks in Experiments 1 and 2 may have produced null results because of the use of Lego stimuli. We therefore conducted an implicit occlusion task using the original DPT avatars, to establish whether the effect found in Experiment 4 is best explained by directional orienting or perspective-taking, and whether the null results in the implicit tasks of Experiments 1 and 2 can be attributed to the stimuli used.

Experiment 5: occlusion task with original stimuli

Materials and methods

The original DPT stimuli were edited to create barriers in the same positions as in the Lego stimuli (see Figure 12). Because the new scene layout required dots to be displayed in positions other than on a flat wall, floating red orbs were used instead of the red discs used in the original DPT. A colour picking tool was used to create the colour for the orbs, and shadows were added to create depth. They were positioned within the eyeline of the avatars, in the same positions as in the Lego stimuli.

Because this task differed from Experiment 2 only in images used, and in the position of the fixation crosses and

pretrial instructions, it was not preregistered separately, as all other details of the preregistration were the same.

Participants. We used the same sample size as in Experiments 1–4. Thirty participants were recruited through the University of Edinburgh Student and Graduate Employment Service. They were compensated with £4 for their participation, which lasted approximately 30 min. Data were excluded from one participant whose task was interrupted by disconnection of the response box, and one participant who had participated in an earlier DPT. Data from these two participants were replaced by new participants. Participants gave written consent, including consent for anonymised data to be shared publicly. Ethical approval was granted by the University of Edinburgh's PPLSREC, reference number 188-1718/2.

Procedure. This task used the same procedure and task design as the implicit condition in Experiment 1.

Results

We applied the data exclusions and transformations described in Experiment 1, removing timed-out trials (0.40%, $n=31$) and erroneous responses (2.51%, $n=192$). There were no responses below 100 ms.

Following the analysis used in Experiments 1 and 2, data limited to the three relevant comparisons (*Behind avatar* vs. *Avatar sees*, *Avatar faces [central]* vs. *Avatar sees*, and *Avatar faces [peripheral]* vs. *Behind avatar*) were modelled using three models identical to those used in Experiment 2.

As in Experiments 1 and 2, and as predicted, the models showed a significant difference between *Avatar sees* and *Behind avatar* (23.88 ms), and no significant difference between *Avatar faces (central)* and *Avatar sees* trials (−0.76 ms, see Table 8). However, contrary to our predictions, there was also no significant difference between *Avatar faces (peripheral)* and *Behind avatar* trials, at 3.31 ms (Figure 13). This suggests that there was no directional orienting effect in this task, and that the difference between *Avatar sees* and *Behind avatar* trials was driven by the spatial distribution of the balls.

Discussion

These results do not support the hypothesis that directional orienting played any role in this implicit task. This continues to conflict with findings of altercentric effects in implicit tasks (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Samson et al., 2010; Santiesteban et al., 2014), and is difficult to explain. One important difference between Experiment 5 and other implicit tasks is the visual complexity of the scene: where other implicit tasks have used an avatar in a consistent position within the scene, we have used two avatars in two possible positions; and where

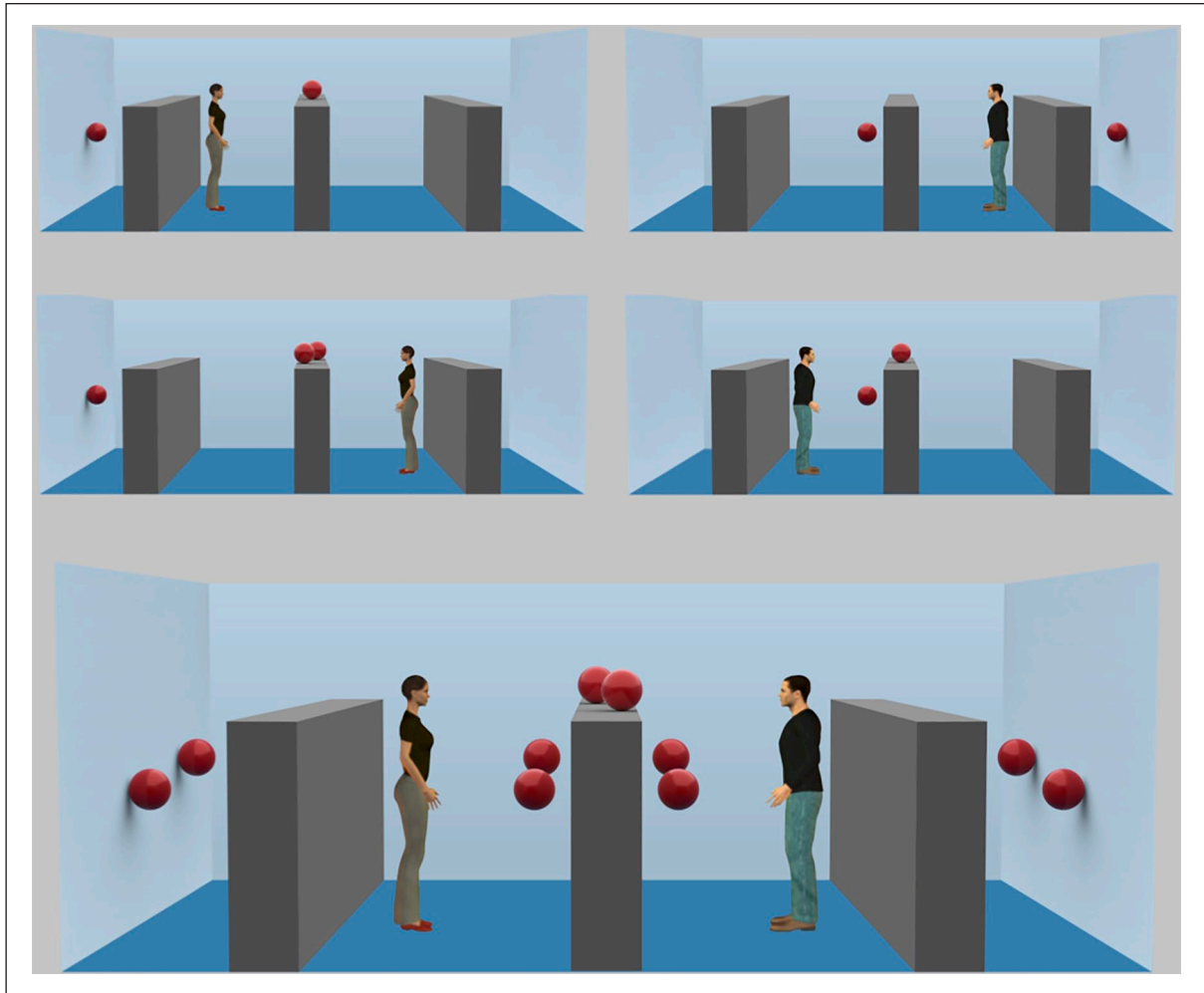


Figure 12. Occlusion task using avatars from the original DPT. The upper four images show example scenes; note that each scene that participants saw featured a single avatar and a maximum of four balls. The lower image shows both potential placement positions for avatars (left or right of the central table) and all possible ball positions (five possible positions, a maximum of two balls in any one position, and a maximum of four balls per scene).

Table 8. Results of Experiment 5.

Slope	β	SE	χ^2	df	p
Behind avatar versus Avatar sees	.021	0.005	17.97	1	<.001***
Avatar faces (central) versus Avatar sees	-.0007	0.005	0.03	1	.86
Behind avatar versus Avatar faces (peripheral)	.003	0.005	0.4	1	.53

SE: standard error.

***p < .001

previous tasks have used goggles, a telescope, a single barrier, or a pair of barriers (one behind and one in front of the avatar), we have three different barriers in our scene, two of which are in front of the avatar. The addition of the second barrier, and the distance between the avatar and any balls behind this barrier on the periphery of the scene, may be sufficient to prevent directional orienting. The visual complexity of this scene design may, therefore, simply be too high for directional orienting to play a role in participants' comprehension of each image.

It would be instructive to replicate Experiment 1 (explicit, implicit, and uncued conditions) using a range of different stimuli, including a simple screen with a window that may be open or closed (Cole et al., 2016), a scene with more realistic depth in the third dimension but still only one barrier (Baker et al., 2016), pictures of humans or human experimenters (Langton, 2018), and alternative occlusion methods such as opaque goggles (Conway et al., 2017; Furlanetto et al., 2016). Further manipulations such as colour saturation and the spacing of dots may also be

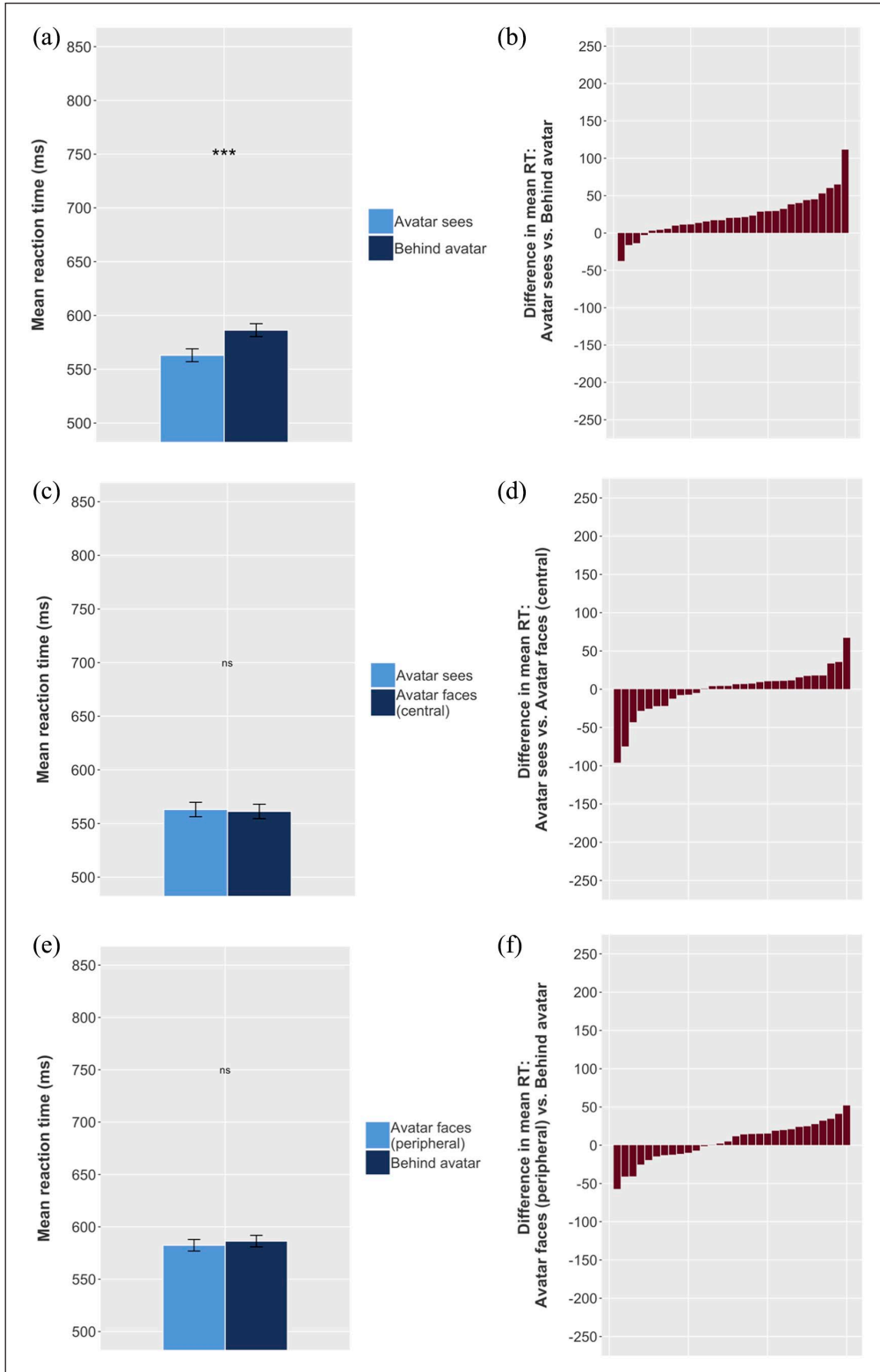


Figure 13. Results of Experiment 5. (a) Mean RT for *Behind avatar* and *Avatar sees* conditions; error bars indicate 95% CIs on the mean of the by-participant means. (b) Each individual participant’s difference between mean *Behind avatar* RT and mean *Avatar sees* RT. Mean RT is higher (i.e., participants respond more slowly) for *Behind avatar* trials (a); a majority of participants show this effect (b). However, there is no difference in RT between (c, d) *Avatar sees* and *Avatar faces (central)* or (e, f) *Behind avatar* and *Avatar faces (peripheral)*.

useful. It is clear that properties of the stimuli affect results in the DPT in a variety of ways, and exploring these effects systematically would greatly clarify the nature and triggering conditions of the altercentric effect. Given the clear range of individual participant differences in responses to the tasks, it may also be the case that much of the DPT literature is underpowered and suffers from sampling error; further research into the individual differences underlying participant responses would be valuable.

The results of Experiment 5 yield no further evidence on whether perspective-taking or directional orienting underlies the altercentric effect found in Experiment 4. It may be the case that the visual complexity of this occlusion task is too high to induce an effect on an implicit task, and that this paradigm is therefore not able to determine whether a consistency effect on a simple implicit task is the result of perspective-taking or directional orienting.

Conclusion and discussion

The results of these five experiments collectively provide evidence that differences in stimuli and task demands, and particularly in perception of the agency and relevance of the on-screen characters, play a substantial role in mediating the results of the DPT. That is, when avatars are more humanoid and realistic, they are more likely to create an altercentric effect, but only in a task of sufficient visual simplicity; and when the avatar's perspective is relevant to the task, it drives a perspective-taking effect.

Experiment 1 showed that uncued tasks (as predicted) do not result in altercentric interference, and that explicit versions of the DPT (Baker et al., 2016; Capozzi et al., 2014; Furlanetto et al., 2016; Marshall et al., 2018; Mattan et al., 2015, 2016; Samson et al., 2010; Wilson et al., 2017) likely do provide evidence of visual perspective-taking, rather than directional orienting. This coheres with our analysis of the literature as containing discrepant findings based on varying implementations of the DPT; namely, that explicit tasks find results consistent with visual perspective-taking rather than directional orienting.

This "visual perspective-taking" could plausibly be achieved by different mechanisms—for instance, by participants spatially representing the dots/discs that are visible from a certain point in the room, regardless of what occupies this position, or by representing the visual perspective of an on-screen figure. The use of a control condition using non-social stimuli such as arrows, lamps, or cameras in an explicit occlusion task could be useful in distinguishing between these mechanisms. That is, if there is a perspective-taking effect on an explicit task for humanoid stimuli, but not for non-social stimuli, it would suggest that the effect is driven by perspective-taking specific to stimuli that represent a human-like perspective. If, however, a perspective-taking effect is found regardless of stimulus type,

this would suggest a spatial representation effect. It is important to note, though, that on-screen avatars have no perspective to represent (they are avatars, not agents), and so perhaps it should be expected that avatars and non-social stimuli would show similar results. It is also possible that spatial representation may be the primary mechanism by which visual perspective-taking is achieved. This would be a fruitful avenue for further research.

This visual perspective-taking is not purely stimulus-driven, instead requiring that participants are motivated to take the perspective of the avatars throughout the task. Given this continuous perspective-taking, it seems that participants maintain awareness of the avatar's perspective (which is relevant on some scenes) throughout the experiment (even on scenes where it is not relevant), and therefore use the avatar's perspective as a cue throughout the task. Mean RT for the explicit condition (720.91 ms) was higher than implicit (594.37 ms) or uncued (578.00 ms) conditions; the experiment was not powered to determine whether this between-subjects difference was statistically significant, but confirmatory research analysing this would be informative, as slower responses on an explicit task could indicate that holding the avatar's perspective in working memory is somewhat effortful. The evidence from Experiment 1 suggests that visual perspective-taking should not be considered automatic, but rather spontaneous, occurring only when relevant; but may still occur involuntarily and rapidly, on trials where it is not necessary for the immediate task (recall that all of our analyses are conducted on trials where participants are only required to take their own perspective).

Although we predicted that the implicit task in Experiment 1 would show directional orienting effects, our results in Experiments 1–3 and 5 failed to match previous findings of directional orienting in implicit versions of the DPT (Cole et al., 2016; Conway et al., 2017; Langton, 2018; Santiesteban et al., 2014). Experiments 2 and 3 investigated whether this could be attributed to (failure to) draw attention to the avatars by placement of the fixation cross and pretrial instructions, or by the greater scene complexity of the stimuli with multiple barriers, two avatars in different positions, and up to four balls. In Experiment 2 we used the fixation cross and instructions to draw attention to the avatar in an implicit task, and still found no evidence of an altercentric effect consistent with the avatar driving directional orienting; the only effect present was better explained by the spatial distribution of the scene. Likewise, in Experiment 3 we simplified our scenes and still found no altercentric effect in an implicit task. However, in Experiment 4, an implicit task using the original DPT stimuli and otherwise identical to Experiment 3 did find an altercentric effect, suggesting an (unanticipated) sensitivity of implicit tasks to the details of the on-screen characters (i.e., cartoon stimuli yield interference, Lego characters do not).

Because of the simplified stimuli, it is not possible to determine whether the altercentric effect found in Experiment 4 was a result of perspective-taking or directional orienting. We therefore conducted an occlusion task using the original avatar stimuli, and otherwise identical to the implicit task in Experiment 1. This task found no evidence of directional orienting (or perspective-taking), with the only effect best explained by the spatial distribution of the scene. This battery of experiments therefore did not confirm one of our main predictions, which was that implicit occlusion tasks would produce an altercentric effect consistent with directional orienting. The greatly increased visual complexity of Experiment 5 stimuli compared with previous implicit occlusion tasks (Cole et al., 2016; Conway et al., 2017; Langton, 2018) may explain why we did not find a directional orienting effect. The unexpected results for the battery of implicit tasks presented in this article suggest the need for future research exploring the variety of ways in which DPT stimuli may affect the results, and the theoretical implications of these variations.

Collectively, these five experiments point to a complex picture of visual perspective-taking, as something occurring spontaneously in dynamic reaction to the immediate environment, based on attentional cues. Our Experiment 1 provides evidence that explicit versions of the DPT likely do provide evidence of visual perspective-taking, rather than directional orienting. The contrast between explicit and implicit/uncued conditions suggests that visual perspective-taking is not purely stimulus-driven, instead requiring that participants are motivated to take the perspective of the avatars throughout the task. Visual perspective-taking should therefore not be considered automatic, but rather spontaneous, occurring only when relevant.

Our results across all five experiments also suggest that the visual complexity of the scene and the perceived agency of the stimuli play a role in driving the appearance of an altercentric effect, contributing further evidence that directional orienting is not automatic, and is instead potentially dependent on the directional cues provided by the stimulus, or on cues to consider the agent's perspective as relevant (albeit not sufficiently to sustain throughout the task, as in an explicit task). Given this result, we emphasise that a clear distinction should be made between processes that are automatic and processes that are spontaneous—that is, not automatic but still rapid, involuntary, and unconscious, arising when necessary, as prompted by the attentional system.

The possibility that perspective taking might be spontaneous raises an important theoretical issue. Specifically, it raises the possibility that directional orienting and perspective taking are in fact not cognitively distinct alternatives. Instead there might be more of a continuum between them, by which directional orienting is a possible input to perspective-taking, with its effect modulated by attention. This possibility is an important topic for future research,

both theoretical and empirical. Finally, we note that while the DPT is proving a fruitful method for exploring questions regarding visual perspective-taking, our results suggest that caution is required to interpret results from a range of tasks with widely varying stimuli and implementation. Given the application of this task to broader questions about theory of mind (Drayton et al., 2018; Yue et al., 2017), it is essential to clarify the precise causes of altercentric interference before using this task to establish group differences in, or the presence or absence of, perspective-taking.

Acknowledgements

The authors thank their colleagues Simon Kirby and Jennifer Culbertson, who provided insight and expertise that greatly assisted this research, and Rachel Kindellan, for her assistance in collecting data for Experiment 5.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 681942. T.S.P. was financially supported by Durham University's Addison Wheeler bequest and by the ERC, under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 609819 [Somics project]. C.O.G. was financially supported by the University of Edinburgh's Principal's Career Development Scholarship, Global Research Scholarship, and Research Support Grants.

ORCID iD

Cathleen O'Grady  <https://orcid.org/0000-0001-5583-0181>

Data Accessibility Statement



The data and materials from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/za3qd>.

Supplemental Material

The Supplementary Material is available at qjep.sagepub.com

Notes

1. It is of course worth noting that the avatars in the dot perspective task (DPT) do not have agency themselves, since they are cartoon-like representations of people rather than actual people. But avatars do at least aim to imitate things that do have agency, unlike, for example, arrows.

2. Materials, experiment code, data, and analysis scripts for all experiments reported in this article are available at <https://osf.io/za3qd/>.
3. Note that this is a technical deviation from the preregistration, in which we did not explicitly state that we would replace such participants.
4. This step was included in the preregistered analysis script, and is necessitated by the experiment design, but was erroneously omitted from the preregistration free-text description of the analysis.
5. A reviewer suggested an exploratory analysis using only the first half of the trials from the implicit and uncued tasks, to confirm whether diminishing effects in the implicit and uncued tasks may explain the results reported. This analysis, using the trials only from the first two blocks of the implicit and uncued tasks, produces the same pattern of results as the preregistered analysis reported here, that is, the same comparisons produced similarly significant or null results. One minor exception is noted in the results of Model 3. The script for this exploratory analysis is available at osf.io/za3qd/.
6. Model syntax: $\log\text{RT} \sim \text{Condition} \times \text{Consistency} + (1 + \text{Consistency} | \text{Participant}) + (1 | \text{Image})$.
7. The exploratory analysis mentioned above using trials from only the first two blocks in the implicit and uncued conditions showed an interaction between Condition and Consistency, $p = .03$.

References

- Apperly, I. (2011). *Mindreaders: The cognitive basis of "theory of mind."* Psychology Press.
- Baayen, H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28. <https://doi.org/10.21500/20112084.807>
- Baker, L. J., Levin, D. T., & Saylor, M. M. (2016). The extent of default visual perspective taking in complex layouts. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 508–516. <https://doi.org/10.1037/xhp0000164>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual Cognition, 23*(8), 1020–1042. <https://doi.org/10.1080/13506285.2015.1132804>
- Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *PLOS ONE, 9*(12), e114210. <https://doi.org/10.1371/journal.pone.0114210>
- Carruthers, P. (2017). Mindreading in adults: Evaluating two systems views. *Synthese, 194*(3), 673–688. <https://doi.org/10.1007/s11229-015-0792-3>
- Cole, G., Atkinson, M., D'Souza, A., & Smith, D. (2017). Spontaneous perspective taking in humans? *Vision, 1*(2), 17. <https://doi.org/10.3390/vision1020017>
- Cole, G., Atkinson, M., Le, A. T., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica, 164*, 165–168. <https://doi.org/10.1016/j.actpsy.2016.01.007>
- Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 454–465. <https://doi.org/10.1037/xhp0000319>
- Drayton, L. A., Santos, L. R., & Baskin-Sommers, A. (2018). Psychopaths fail to automatically take the perspective of others. *Proceedings of the National Academy of Sciences, 115*(13), 3302–3307. <https://doi.org/10.1073/pnas.1721903115>
- Freundlieb, M., Kovács, A. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance, 42*(3), 401–412. <https://doi.org/10.1037/xhp0000153>
- Freundlieb, M., Kovács, A. M., & Sebanz, N. (2018). Reading your mind while you are reading—Evidence for spontaneous visuospatial perspective taking during a Semantic Categorization Task. *Psychological Science, 29*(4), 614–622. <https://doi.org/10.1177/0956797617740973>
- Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 158–163. <https://doi.org/10.1037/xhp0000138>
- Gardner, M. R., Bileviciute, A., & Edmonds, C. (2018a). Implicit mentalising during level-1 visual perspective-taking indicated by dissociation with attention orienting. *Vision, 2*(1), 3. <https://doi.org/10.3390/vision2010003>
- Gardner, M. R., Hull, Z., Taylor, D., & Edmonds, C. J. (2018b). "Spontaneous" visual perspective-taking mediated by attention orienting that is voluntary and not reflexive. *Quarterly Journal of Experimental Psychology, 71*(4), 1020–1029. <https://doi.org/10.1080/17470218.2017.1307868>
- Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9*(2), 131–143. <https://doi.org/10.1177/1745691613518076>
- Langton, S. (2018). I don't see it your way: The dot perspective task does not gauge spontaneous perspective taking. *Vision, 2*(1), 6. <https://doi.org/10.3390/vision2010006>
- LaPointe, M. R. P., Cullen, R., Baltaretu, B., Campos, M., Michalski, N., Sri Satgunarajah, S., Cadieux, M. L., Pachai, M. V., & Shore, D. I. (2016). An attentional bias for LEGO® people using a change detection task: Are LEGO® people animate? *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 70*(3), 219–231. <https://doi.org/10.1037/cep0000077>
- Marotta, A., Lupiáñez, J., Martella, D., & Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: Two qualitatively different modes of attentional selection. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 326–335. <https://doi.org/10.1037/a0023959>
- Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing?: Results from two

- level-1 perspective-taking tasks. *PLOS ONE*, *13*(3), Article e0194101. <https://doi.org/10.1371/journal.pone.0194101>
- Mattan, B. D., Quinn, K. A., Apperly, I. A., Sui, J., & Rotshtein, P. (2015). Is it always me first? Effects of self-tagging on third-person perspective-taking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1100–1117. <https://doi.org/10.1037/xlm0000078>
- Mattan, B. D., Rotshtein, P., & Quinn, K. A. (2016). Empathy and visual perspective-taking performance. *Cognitive Neuroscience*, *7*(1–4), 170–181. <https://doi.org/10.1080/17588928.2015.1085372>
- Michael, J., Wolf, T., Letesson, C., Butterfill, S., Skewes, J., & Hohwy, J. (2018). Seeing it both ways: Using a double-cuing task to investigate the role of spatial cuing in Level-1 visual perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 693–702. <https://doi.org/10.1037/xhp0000486>
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. <https://doi.org/10.1037/0033-2909.132.2.297>
- Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *Quarterly Journal of Experimental Psychology*, *68*(10), 1931–1951. <https://doi.org/10.1080/17470218.2015.1023206>
- Peirce, J. (2010). *PsychoPy—Psychology software for Python*.
- Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236. <https://doi.org/10.1016/j.cognition.2010.08.003>
- R Core Team. (2015). *R: A language and environment for statistical computing*.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266. <https://doi.org/10.1037/a0018729>
- San Juan, V., & Astington, J. W. (2017). Does language matter for implicit theory of mind? The effects of epistemic verb training on implicit and explicit false-belief understanding. *Cognitive Development*, *41*, 19–32. <https://doi.org/10.1016/j.cogdev.2016.12.003>
- Santesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–937. <https://doi.org/10.1037/a0035175>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2017). *Afex: Analysis of factorial experiments. R package version 0.17-8*.
- Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52. <https://doi.org/10.1016/j.cognition.2016.01.014>
- Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults: egocentrism and automatic perspective taking. *Child Development*, *83*, 452–460. <https://doi.org/10.1111/j.1467-8624.2011.01730.x>
- Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80–102.
- Westra, E. (2017). Spontaneous mindreading: A problem for the two-systems account. *Synthese*, *194*(11), 4559–4581. <https://doi.org/10.1007/s11229-016-1159-0>
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482. <https://doi.org/10.1007/BF03395630>
- Wilson, C. J., Soranzo, A., & Bertamini, M. (2017). Attentional interference is modulated by salience not sentience. *Acta Psychologica*, *178*, 56–65. <https://doi.org/10.1016/j.actpsy.2017.05.010>
- Yue, T., Jiang, Y., Yue, C., & Huang, X. (2017). Differential effects of oxytocin on visual perspective taking for men and women. *Frontiers in Behavioral Neuroscience*, *11*, Article 228. <https://doi.org/10.3389/fnbeh.2017.00228>
- Zhao, X., Cusimano, C., & Malle, B. F. (2015). In search of triggering conditions for spontaneous visual perspective taking. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2811–2816). Cognitive Science Society.